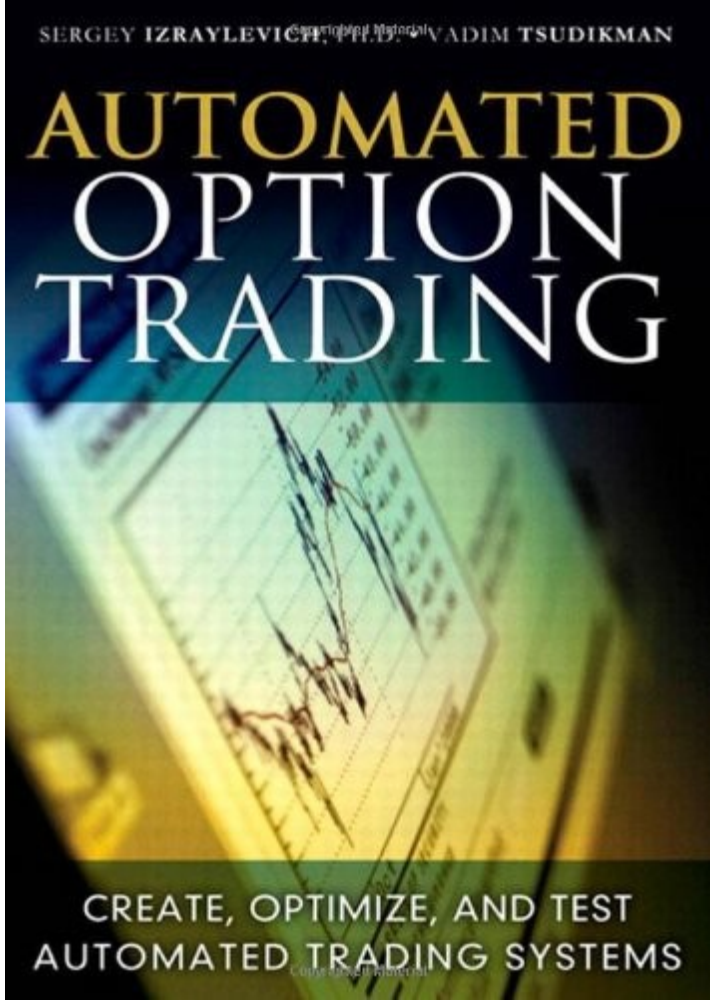
SERGEY IZRAYLEVICH, PH.D. VADIM TSUDIKMAN
AUTOMATED
OPTION
TRADING
CREATE, OPTIMIZE, AND TEST
AUTOMATED TRADING SYSTEMS

VISIT...

SERGEY IZRAYLEVICH, PH.D. • VADIM TSUDIKMAN

AUTOMATED OPTION TRADING

CREATE, OPTIMIZE, AND TEST AUTOMATED TRADING SYSTEMS

Automated Option Trading

Create, Optimize, and Test Automated Trading Systems

Sergey Izraylevich, Ph.D., and
Vadim Tsudikman

Vice President, Publisher: Tim Moore
Associate Publisher and Director of Marketing: Amy Neidlinger
Executive Editor: Jim Boyd
Editorial Assistant: Pamela Boland
Operations Specialist: Jodi Kemper
Senior Marketing Manager: Julie Phifer
Assistant Marketing Manager: Megan Graue
Cover Designer: Alan Clements
Managing Editor: Kristy Hart
Project Editor: Betsy Harris
Copy Editor: Cheri Clark
Proofreader: Debbie Williams
Senior Indexer: Cheryl Lenser
Senior Compositor: Gloria Schurick
Manufacturing Buyer: Dan Uhrig

© 2012 by Pearson Education, Inc.
Publishing as FT Press
Upper Saddle River, New Jersey 07458

This book is sold with the understanding that neither the author nor the publisher is engaged in rendering legal, accounting, or other professional services or advice by publishing this book. Each individual situation is unique. Thus, if legal or financial advice or other expert assistance is required in a specific situation, the services of a competent professional should be sought to ensure that the situation has been evaluated carefully and appropriately. The author and the publisher disclaim any liability, loss, or risk resulting directly or indirectly, from the use or application of any of the contents of this book.

FT Press offers excellent discounts on this book when ordered in quantity for bulk purchases or special sales. For more information, please contact U.S. Corporate and Government Sales, 1-800-382-3419, corpsales@pearsontechgroup.com. For sales outside the U.S., please contact International Sales at international@pearson.com.

Company and product names mentioned herein are the trademarks or registered trademarks of their respective owners.

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

Printed in the United States of America

First Printing March 2012

ISBN-10: 0-13-247866-8
ISBN-13: 978-0-13-247866-3

Pearson Education LTD.
Pearson Education Australia PTY, Limited
Pearson Education Singapore, Pte. Ltd.
Pearson Education Asia, Ltd.
Pearson Education Canada, Ltd.
Pearson Educatión de Mexico, S.A. de C.V.
Pearson Education—Japan
Pearson Education Malaysia, Pte. Ltd.

Library of Congress Cataloging-in-Publication Data

Izraylevich, Sergey, 1966-

Automated option trading : create, optimize, and test automated trading systems / Sergey Izraylevich, Vadim Tsudikman.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-13-247866-3 (hardcover : alk. paper)
1. Stock options. 2. Options (Finance) 3. Electronic trading of securities. 4. Investment analysis. 5. Portfolio management. I. Tsudikman, Vadim, 1965- II.
Title.
HG6042.I968 2012
332.64’53—dc23

2011048038

This book is dedicated to our parents,
Izraylevich Olga, Izraylevich Vladimir,
Tsudikman Rachel, Tsudikman Jacob.

Contents

Introduction

Chapter 1. Development of Trading Strategies

1.1. Distinctive Features of Option Trading Strategies

1.1.1. Nonlinearity and Options Evaluation

1.1.2. Limited Period of Options Life

1.1.3. Diversity of Options

1.2. Market-Neutral Option Trading Strategies

1.2.1. Basic Market-Neutral Strategy

1.2.2. Points and Boundaries of Delta-Neutrality

1.2.3. Analysis of Delta-Neutrality Boundaries

1.2.4. Quantitative Characteristics of Delta-Neutrality Boundaries

1.2.5. Analysis of the Portfolio Structure

1.3. Partially Directional Strategies

1.3.1. Specific Features of Partially Directional Strategies

1.3.2. Embedding the Forecast into the Strategy Structure

1.3.3. The Call-to-Put Ratio at the Portfolio Level

1.3.4. Basic Partially Directional Strategy

1.3.5. Factors Influencing the Call-to-Put Ratio in an Options Portfolio

1.3.6. The Concept of Delta-Neutrality as Applied to a Partially Directional Strategy

1.3.7. Analysis of the Portfolio Structure

1.4. Delta-Neutral Portfolio as a Basis for the Option Trading Strategy

1.4.1. Structure and Properties of Portfolios Situated at Delta-Neutrality Boundaries

1.4.2. Selection of an Optimal Delta-

Neutral Portfolio

Chapter 2. Optimization

2.1. General Overview

2.1.1. Parametric Optimization

2.1.2. Optimization Space

2.1.3. Objective Function

2.2. Optimization Space of the Delta-Neutral Strategy

2.2.1. Dimensionality of Optimization

2.2.2. Acceptable Range of Parameter Values

2.2.3. Optimization Step

2.3. Objective Functions and Their Application

2.3.1. Optimization Spaces of Different Objective Functions

2.3.2. Interrelationships of Objective Functions

2.4. Multicriteria Optimization

2.4.1. Convolution

2.4.2. Optimization Using the Pareto Method

2.5. Selection of the Optimal Solution on the Basis of Robustness

2.5.1. Averaging the Adjacent Cells

2.5.2. Ratio of Mean to Standard Error

2.5.3. Surface Geometry

2.6. Steadiness of Optimization Space

2.6.1. Steadiness Relative to Fixed Parameters

2.6.2. Steadiness Relative to Structural Changes

2.6.3. Steadiness Relative to the Optimization Period

2.7. Optimization Methods

2.7.1. A Review of the Key Direct Search

Methods

2.7.2. Comparison of the Effectiveness of Direct Search Methods

2.7.3. Random Search

2.8. Establishing the Optimization Framework: Challenges and Compromises

Chapter 3. Risk Management

3.1. Payoff Function and Specifics of Risk Evaluation

3.1.1. Linear Financial Instruments

3.1.2. Options as Nonlinear Financial Instruments

3.2. Risk Indicators

3.2.1. Value at Risk (VaR)

3.2.2. Index Delta

3.2.3. Asymmetry Coefficient

3.2.4. Loss Probability

3.3. Interrelationships Between Risk Indicators

3.3.1. Method for Testing the Interrelationships

3.3.2. Correlation Analysis

3.4. Establishing the Risk Management System: Challenges and Compromises

Chapter 4. Capital Allocation and Portfolio Construction

4.1. Classical Portfolio Theory and Its Applicability to Options

4.1.1. Classical Approach to Portfolio Construction

4.1.2. Specific Features of Option Portfolios

4.2. Principles of Option Portfolio Construction

4.2.1. Dimensionality of the Evaluation System

4.2.2. Evaluation Level

4.3. Indicators Used for Capital Allocation

4.3.1. Indicators Unrelated to Return and Risk Evaluation

4.3.2. Indicators Related to Return and Risk Evaluation

4.4. One-Dimensional System of Capital Allocation

4.4.1. Factors Influencing Capital Allocation

4.4.2. Measuring the Capital Concentration in the Portfolio

4.4.3. Transformation of the Weight Function

4.5. Multidimensional Capital Allocation System

4.5.1. Method of Multidimensional System Application

4.5.2. Comparison of Multidimensional and One-Dimensional Systems

4.6. Portfolio System of Capital Allocation

4.6.1. Specific Features of the Portfolio System

4.6.2. Comparison of Portfolio and Elemental Systems

4.7. Establishing the Capital Allocation System: Challenges and Compromises

Chapter 5. Backtesting of Option Trading Strategies

5.1. Database

5.1.1. Data Vendors

5.1.2. Database Structure

5.1.3. Data Access

5.1.4. Recurrent Calculations

5.1.5. Checking Data Reliability and Validity

5.2. Position Opening and Closing Signals

5.2.1. Signals Generation Principles

5.2.2. Development and Evaluation of Functionals

5.2.3. Filtration of Signals

5.3. Modeling of Order Execution

5.3.1. Volume Modeling

5.3.2. Price Modeling

5.3.3. Commissions

5.4. Backtesting Framework

5.4.1. In-Sample Optimization and Out-of-Sample Testing

5.4.2. Adaptive Optimization

5.4.3. Overfitting Problem

5.5. Evaluation of Performance

5.5.1. Single Event and Unit of Time Frame

5.5.2. Review of Strategy Performance Indicators

5.5.3. The Example of Option Strategy Backtesting

5.6. Establishing the Backtesting System: Challenges and Compromises

Bibliography

Appendix

Basic Notions

Payoff Functions

Separate Options

Option Combinations

Index

Acknowledgments

The authors would like to express their gratitude to the team of High Technology Invest Inc. Special thanks are due to Arsen Balishyan, Ph.D., CFA, and Vladislav Leonov, Ph.D., and Eugen Masherov, Ph.D., for their continued help at all stages of this project.

About the Authors

Sergey Izraylevich, Ph.D., Chairman of the Board of High Technology Invest Inc., has traded options for well over a decade and currently creates automated systems for algorithmic option trading. A *Futures* magazine columnist, he has authored numerous articles for highly rated, peer-reviewed scientific journals. He began his career as a lecturer at The Hebrew University of Jerusalem and Tel-Hay Academic College, receiving numerous awards for academic excellence, including Golda Meir's Prize and the Max Shlomiok honor award of distinction.

Vadim Tsudikman, President of High Technology Invest Inc., is a financial consultant and investment advisor specializing in derivatives valuation, hedging, and capital allocation in extreme market environments. With more than 15 years of option trading experience, he develops complex trading systems based on multicriteria analysis and genetic optimization algorithms.

Izraylevich and Tsudikman coauthored *Systematic Options Trading* (FT Press) and regularly coauthor *Futures* magazine articles on cutting-edge issues related to option pricing, volatility, and risk management.

Introduction

This book presents a concept of developing an automated system tailored specifically for options trading. It was written to provide a framework for transforming investment ideas into properly defined and formalized algorithms allowing consistent and disciplined realization of testable trading strategies.

Extensive literature has been published in the past decades regarding systematic, algorithmic, automated, and mechanical trading. In the Bibliography of this book, we list some of the comprehensive works that deserve special attention. However, all books dedicated to the creation of automated trading systems deal with traditional investment tools, such as stocks, futures, or currencies. Although the development of options-oriented systems requires accounting for numerous specific features peculiar to these instruments, automated trading of options remains beyond the scope of professional literature. The philosophy, logic, and quantitative procedures used in the creation of automated systems for options trading are completely different from those used in conventional trading algorithms. In fact, all the components of a system intended for automated trading of options (strategy development, optimization, capital allocation, risk management, backtesting, performance measurement) differ significantly from their analogs in the systems intended for trading of plain assets. This book describes consecutively the key stages of creating automated systems intended specifically for options trading.

Automated trading of options represents a continuous process of valuation, structuring, and long-term management of investment portfolios (rather than individual instruments). Due to the nonlinearity of options, the expected returns and risks of their complex portfolios cannot be estimated by simple summation of characteristics corresponding to individual options. Special approaches are required to evaluate portfolios containing options (and their combinations) related to different underlying assets. In this book we discuss such approaches, describe systematically the core properties of option portfolios, and consider the specific features of automated options trading at the portfolio level.

The Book Structure

An automated trading system represents a package of software modules performing the functions of developing, formalizing, setting up, and testing trading strategies.

Chapter 1, "Development of Trading Strategies," discusses the development and formalization of option strategies. Since there is a huge multitude of trading strategies somehow related to options, we limit our discussion to market-neutral strategies. The reason for selecting this particular type of option strategies relates to its wide popularity among private and institutional investors.

Strategy setup includes optimization of its parameters, capital allocation between portfolio elements, and risk management. Chapter 2, "Optimization," deals with various optimization aspects. In this chapter we discuss various properties of optimization spaces, different types of objective functions and their interrelationships, several methods of multicriteria optimization, and problems of optimization steadiness relative to small changes in the parameters and strategy structure. Special attention is given to the application of traditional methods of parametric optimization to complex option portfolios.

In Chapter 3, "Risk Management," we discuss a set of option-specific risk indicators that can be used for developing a multicriteria risk management system. We investigate the influence of different factors on the effectiveness of the risk indicator and on the number of indicators needed for effective risk measuring.

In Chapter 4, "Capital Allocation and Portfolio Construction," we consider various aspects of capital allocation among the elements of an option portfolio. Capital can be allocated on the basis of different indicators not necessarily expressing return and risk. This chapter describes the step-by-step process of constructing a complex portfolio out of separate option combinations.

The testing of option strategies using historical data is discussed in Chapter 5, "Backtesting of Option Trading Strategies." In this chapter we stress the particularities of creating and maintaining option databases and provide methods to verify data accuracy and reliability. The problem of overfitting and the main approaches to solving it are also discussed. Performance evaluation of option strategies is also the topic of this chapter.

Strategies Considered in This Book

The nature of options makes it possible to create a considerable number of speculative trading strategies. Those can be based on different approaches encompassing the variety of principles and valuation techniques.

In many strategies options are used as auxiliary instruments to *hedge main positions*. In this book we are not going to delve into this field of options application since hedging represents only one constituent part

of such trading strategies, but not their backbone.

Options may be used to create *synthetic underlying positions.* In this case the investor aims for the payoff profiles of an option combination and its underlying asset to coincide. This can increase trading leverage significantly. However, apart from leverage, automated trading of synthetic assets is no different from trading in underlying assets (besides the certain specificity regarding execution of trading orders, higher brokerage commissions, and the necessity to roll over positions). Thus, we will not dwell on such strategies either.

Most trading strategies dealing with plain assets (not options) are based on the *forecast of the direction of their price movement* (we will call them *directional strategies*). Options can also be used in such strategies. For example, different kinds of option combinations, commonly referred to as *spreads,* benefit from the increase in the underlying price (bull spreads), or from its decline (bear spread). Despite the fact that trading strategies based on such option combinations possess many features distinguishing them from plain assets strategies, the main determinant of their performance is the accuracy of price forecasts. This quality makes such strategies quite similar to common directional strategies, and therefore we will not consider them in this book.

The focus of this book is on strategies that exploit the specific features of options. One of the key differences of options from other investment assets is the nonlinearity of their payoff functions. In the trading of stocks, commodities, currencies, and other linear assets, all profits and losses are directly proportional to their prices. In the case of options, however, position profitability depends not only on the direction of the price movement, but on many other factors as well. Combining different options on the same underlying asset can bring about almost any form of the payoff function.

This feature of options permits the creation of positions that depend not only on the direction and the amplitude of price fluctuations, but also on many other parameters, including volatility, time left until the expiration, and so forth. The main subject of our consideration is a special type of trading strategies sharing one common property referred to as *market-neutrality*. With regard to options, market-neutrality means that (1) small changes in the underlying price do not lead to a significant change in the position value, and (2) given larger price movements, the position value changes by approximately the same amount regardless of the direction of the underlying price movement. In reality these rules do not always hold, but they serve as a general guideline for a trader striving for market-neutrality. The main analytical instrument used to create market-neutral positions is

delta. The position is market-neutral if the sum of the deltas of all its components (options and underlying assets) is equal to or close to zero. Such positions are referred to as *delta-neutral*.

Another type of trading strategy that will be considered in this book is a set of market-neutral strategies whose algorithms contain certain directional elements. Although in this case positions are created while taking into consideration the value of delta, its reduction to zero is not an obligatory requirement. Forecasts of the directions of future price movements represent an integral part of such a strategy. These forecasts can be incorporated into the strategy structure in the form of biased probability distributions or asymmetrical option combinations, or by application of technical and fundamental indicators. We will call such strategies *partially directional*.

Generally, automated strategies are designed to trade one or just a few financial instruments (mainly futures on a given underlying asset). Even if several instruments are traded simultaneously, in most cases positions are opened, closed, and analyzed independently. Options are no exception. Most traders develop systems oriented solely at trading OEX (options on S&P 100 futures) or options on oil futures. In this book we will consider strategies intended to trade an *unlimited number of options relating to many underlying assets*. All positions created within one trading strategy will be *evaluated and analyzed jointly as a whole portfolio*.

Scientific and Empirical Approaches to Developing Automated Trading Strategies

There are two main approaches to the development of automated trading strategies. The first approach is based on the principles and concepts defined by a strategy developer. All the elements composing such a strategy originate from economic knowledge, fundamental estimates, expert opinions, and so forth. Formalization of such knowledge, estimates, and assumptions in the form of algorithmic rules provides a basis for creating an automated trading strategy. Following the example of Robert Pardo, we will call this a *scientific approach*.

At its extreme, the scientific approach provides for a total rejection of optimization procedures. All the rules and parameters of a trading system are determined solely on the basis of knowledge and forecasts of the developer. Apparently, the likelihood of creating a profitable strategy, while avoiding engagement in optimization procedures, is extremely low. Scientific approach in its pure form is hardly applicable in real trading.

The alternative approach is based on the complete denial of any a priori established theories, models, and principles while developing automated trading strategies. This approach requires extensive use of computer technologies to search for profitable trading rules. All algorithms can be tested for this task (with no concern for any economically sound reasons standing behind their application). Candidate algorithms can be selected from a number of ready-made alternatives available or actually constructed by the system developer. The method of algorithm creation is not determined by preliminary assumptions and is not limited by any exogenous reasoning. Trading rules are selected solely on the basis of their testing using historical data. The resulting strategy is devoid of any behavioral logic or economic sense. Following Robert Pardo, we will call it an *empirical approach*.

At its extreme, the empirical approach is a purposeful quest for algorithms and parameters that maximize simulated profit (minimize loss or satisfy any other utility function). This approach is based exclusively on optimization. Nowadays there is a wide choice of high-technology software that facilitates fast development of effective algorithms and provides for establishment of optimal parameter sets. For example, neuron networks and genetic methods represent powerful tools that enable relatively fast finding of optimal solutions through the creation of self-learning systems.

Usually trading strategies constructed on the basis of the empirical approach show remarkable results when tested on historical time series, but demonstrate failure in real trading. The reason for this is overfitting. Even walk-forward analysis does not eliminate this threat because the significant number of degrees of freedom (which is not unusual under the empirical approach) allows choosing such a set of trading rules and parameters that would generate satisfactory results not only during the optimization period, but in the walk-forward analysis as well (we will examine this in detail in Chapter 5). Thus, practical use of the empirical approach exclusively is risky and hardly applicable in real trading.

Rational Approach to Developing Automated Trading Strategies

Most developers of automated trading strategies combine scientific methods and empirical approaches. On the one hand, strategies resulting from such combining are based on strong economic grounds. On the other hand, they benefit from the numerous advantages of optimization and from the impressive progress in computer intelligence. We will call this the *rational approach*.

Under the rational approach a set of rules, determining the general structure of a trading strategy, is formed at the initial stage of strategy creation. These rules are based on the prior knowledge and assumptions about market behavior. The results of statistical research, either received by the strategy developer or obtained from scientific publications and private sources, can also be used to shape the general framework. Obviously, patterns established during such research introduce certain logic into the strategy under development. At the same time, statistical research may result in the discovery of inexplicable relationships lacking any economic sense behind them. Such relationships should be treated with special care since they may either be random in nature or result from data mining.

The initial stage of strategy creation is based mainly on the elements of scientific approach. At this stage the following must be determined:

- Principles of generating the signals for opening and closing trading positions
- Indicators used to generate open and close signals
- A universe of investment assets that are both available and suitable for trading
- Requirements to the portfolio and restrictions imposed on it
- Capital allocation among different portfolio elements
- Methods and instruments of risk management

At the next stage of developing a trading strategy, the rules laid down on the basis of scientific approach are formalized in the form of computable algorithms. This stage is congested with elements of the empirical approach. These are the essential steps:

- Defining specific parameters. All rules formulated on the basis of the scientific approach should be formalized using a certain number of parameters.
- Specifying the algorithms for parameter calculation. Different algorithms may be invented for calculating the same parameter.
- Establishing the procedures for the selection of parameter values. This requires adopting a certain optimization technique.

Usually the decision on the number of parameters and selection of methods for their optimization does not depend on the economic considerations of the developer, but follows from the specific

requirements to the strategy and from its technical constraints. These requirements and constraints are developed with regard to the reliability, stability, and other strategy features, among which the capability to avoid the overfitting is one of the most important properties.

In this book we will follow the principles of rational approach to the creation of trading strategies. The main task of the developer is to combine methods attributed to the scientific and the empirical approaches in a reasonable and balanced manner. In order to accomplish this task successfully, all basic components of a trading strategy should be clearly identified as belonging either to components that are set on the basis of well-founded reasoning or to an alternative category of components that are formed primarily by applying various optimization methods.

Chapter 1. Development of Trading Strategies

1.1. Distinctive Features of Option Trading Strategies

Developing option trading strategies requires taking into account numerous specificities peculiar to these instruments. In this section we consider three main characteristics that distinguish option-based strategies from other trading systems.

1.1.1. Nonlinearity and Options Evaluation

Options, in contrast to other assets, have nonlinear payoff functions. This should be taken into account while evaluating their investment potential and generating trading signals. Most automated strategies oriented toward linear assets trading generate open and close signals using specially designed indicators. Technical gauges assessing price trends, dynamics of trading volume, and overbought/oversold patterns are among the most frequently used tools. Automated trading of linear assets can also be based on fundamental characteristics. Whatever variables are utilized, the quality of trading signals depends on the effectiveness of indicators used to forecast the future price movements.

Option strategies considered in this book do not explicitly require forecasts of price movement direction (although they may be used as auxiliary instruments). That is why trading signals in automated strategies designed for options trading are generated through the application of special criteria (rather than traditional indicators) evaluating potential profitability and risk on the basis of other principles. The major task of the criteria is to identify undervalued and overvalued investment assets.

Fair value of an option is determined by the extent of uncertainty regarding future fluctuations in its underlying asset price. The higher the uncertainty is, the higher the option price is. Strictly speaking, the option price depends on the probability distribution of underlying asset price outcomes. The market estimate of the degree of uncertainty is implied in option prices. Any particular investor can develop his own estimate (by applying the probability theory, or different mathematical and statistical methods, or based on his personal reasoning). If the magnitude of uncertainty estimated by the investor is different from the market's estimate, the investor may assume that the option under consideration is overpriced or underpriced. The interrelationship between these two uncertainties is the main

philosophic principle underlying criteria creation. This interrelationship may be expressed either directly or indirectly, but it always forms the main component of the algorithm of criteria calculation.

Criteria based on the interrelationship between two uncertainties evaluate the fairness of the option market price. The calculation algorithm of a particular criterion must express values of both uncertainties numerically, reduce them to the same dimension, and compare with each other. If their values are equal or close, then the option is fairly valued by the market. If the uncertainty estimated by the developer is significantly higher (lower) than that of the market, options are underpriced (overpriced). The criterion effectiveness depends on its capability to express the relationship between the divergence of two uncertainties and the magnitude of option over- or underpricing. In our previous book (Izraylevich and Tsudikman, 2010) we described algorithms for calculation of many options-specific criteria. We also discussed the main principles of criteria creation, optimization of their parameters, and different approaches to the evaluation of criteria effectiveness.

1.1.2. Limited Period of Options Life

Another distinctive feature of options is that, unlike other financial instruments, derivatives have a limited life span. This imposes certain limitations on the duration of the position holding period and in some cases requires the execution of the rollover procedures (which leads to additional expenses due to slippage and transaction costs).

In plain assets trading each signal for the position opening is followed by the corresponding signal for closing the same position. In the case of options there may be no close signals if the trading strategy holds positions until expiration. When options expire out-of-the-money, the open signals have no corresponding close signals. If options are in-the-money at expiration, the signals for opening option positions will correspond to close signals relating to their underlying assets. Furthermore, open and close signals can have the same direction (that is, both can be either “buy” or “sell”).

Generation of an open signal for any financial instrument implies that the trading system has detected the divergence of analytically estimated fair value of the instrument from its market price. However, this divergence may persist at the market for an infinitely long time. Even if the calculus was executed correctly, the market price of the instrument with an indefinite lifetime may not converge with its estimate during the whole life span of the strategy. Consequently, the developer of the trading strategy will not be able to judge the

accuracy of his valuation algorithms. On the other hand, options have a fixed expiration date. After this date, the developer can draw a definitive conclusion about the accuracy of the fair value estimate. This feature distinguishes options from other assets for which the period of estimates verification cannot be determined objectively.

1.1.3. Diversity of Options

For each underlying asset there is a multitude of options with different strike prices and expiration dates. At any moment in time, some of them can be over- or undervalued. This allows creating a great number of option combinations with long positions for undervalued options and short positions for overvalued ones.

The number of different option contracts that are potentially available for trading can be estimated as

$$\text{options quantity} = \sum_{i=1}^{n} s_i \times e_i,$$

where n is the number of underlying assets, s_i is the number of strike prices for underlying asset i, and e_i is the number of expiration series existing for this underlying. For each underlying asset it is possible to create $3^{s_i \times e_i}$ combinations (assuming that each option may be included in the combination as a short or long position, or may not be included at all). Accordingly, for n underlying assets the number of constructible option combinations is

$$\text{combinations quantity} = \sum_{i=1}^{n} 3^{s_i \times e_i}.$$

Suppose that the scope of a particular trading system is limited to 100 underlying assets. (In fact, the number of assets with exchange-traded options is much higher; the U.S. stock market alone has several thousand stocks with more or less actively traded options.) Assume also that each underlying asset has about ten option contracts (in reality, their number is much higher). Then at each moment the system constructs and evaluates more than six million combinations (even assuming that all options are included in combinations in equal proportions). If an unequal ratio for different options within the same combination is allowed (as is the case in partially directional strategies; see section 1.3), the system creates a really huge number of combinations.

Of course, there is no reason to consider all randomly generated combinations. The developer usually limits his system to those combinations whose payoff functions correspond to the trading

strategy. These potentially appropriate combinations will be further filtered by liquidity, spread, pending corporate events, fundamental characteristics, and many other parameters. Nevertheless, after all filtrations, the initial set of suitable combinations will comprise about a million variants. Such variety is unattainable for stocks, commodities, currencies, or any other nonderivative assets.

1.2. Market-Neutral Option Trading Strategies

Market-neutral positions are relatively insensitive to the direction of price fluctuations. With regard to options, this means that a small change in the price of the underlying asset has no significant effect on position value. If a large price movement does affect the position, its value changes by approximately the same amount regardless of the direction of price movement. The main tool used in creating market-neutral strategies is the index delta. The position is considered to be market-neutral if the cumulative delta of all its components equals zero. Such a position is called "delta-neutral." We will use both terms ("market-neutrality" and "delta-neutrality") interchangeably.

1.2.1. Basic Market-Neutral Strategy

In this section we describe the general and, in many respects, the simplest form of a market-neutral strategy. Considering the basic form of this strategy is useful for understanding its specific properties, for analysis of its main structural elements, and for comparison with other classes of option trading strategies.

Signals for positions opening. Position opening signals are generated according to a single indicator. For the basic strategy, we have chosen the "expected profit on the basis of lognormal distribution" criterion as an indicator. The opening signal is generated if the criterion value calculated for a given option combination exceeds the predetermined threshold value. The range of allowable values for the threshold parameter stretches from zero to infinity. The exact threshold value is determined by optimization. Opening signals are calculated daily and positions are opened for all combinations receiving trading signals.

Signals for positions closing. All positions that have been opened are held until expiration. After the expiration, all positions in underlying assets, resulting from execution of options that have expired in-the-money, are closed at the next trading day.

Indicators used to generate signals. The algorithm for calculating the "expected profit on the basis of lognormal distribution" criterion is described in our previous book (Izraylevich and Tsudikman, 2010).

Computation of this criterion requires two parameters: the expected value of the underlying asset price and the variance of the normal distribution of the stock price logarithm. The former parameter is usually set a priori by the strategy developer according to the principles of scientific approach. As applied to a market-neutral strategy, it would be reasonable to assume that the expected price is equal to the current underlying asset price as of the criterion calculation date. This means that the current price is considered to be the most accurate estimate of the future price. (The alternative approach is to set this parameter by applying some expert forecasts or different instruments of fundamental and technical analyses.) The value of the second parameter is generally calculated as the square of historical volatility of the underlying asset price. This parameter includes an additional subparameter: the length of the historical period used to calculate volatility. In most cases the historical horizon is set empirically through optimization (for the basic strategy we fix it for 120 days).

The universe of investment assets. The initial set of investment assets that are both allowable and available for the basic trading strategy includes all options on stocks in the S&P 500 index. As a unit of the investment object, we set a combination of options (rather than a separate option) relating to a specific underlying asset. The acceptable types of option combinations are limited to long and short strangles and straddles.

Requirements and restrictions. Taking into account the liquidity and slippage risks, the basic strategy is limited to using the strike prices that do not go beyond 50% from the current underlying asset price (that is, the maximum allowable value for the parameter "strikes range" is 50%). For the same reason, we set the maximum allowable value for the time left to options expiration at 120 working days. These restrictions on the ranges of permissible values follow from our prior experience (that is, the scientific approach is used). The exact values for these parameters have to be determined through optimization (that is, by applying the empirical approach methods).

Money management. With regard to the basic strategy, the problem of money management is reduced to distributing funds between risk-free money market assets and the investment portfolio (see Chapter 4, "Capital Allocation and Portfolio Construction," for details). At each moment in time, after new open signals have been generated, the system determines which part of unused capital should be invested in these new positions. For the basic form of the market-neutral strategy, we adopt the simplest rule. The capital is distributed equally among all trading days within the operating cycle of the strategy (operating

cycles are determined by the expiration dates). Within each trading day all funds available at that day are invested.

Capital allocation within the investment portfolio. Distribution of capital between different option combinations is based on the principle of stock equivalency (see Chapter 4 for details). According to this principle, the position volume is determined in such a manner that should the options be exercised, the funds required to fulfill the obligation (that is, to take either a long or a short position in the underlying asset) will be approximately equal for each combination.

Risk management. As it follows from the nature of a market-neutral strategy, the main guideline of risk management is adherence to the principle of delta-neutrality. Consequently, the main instrument of the risk management for the basic strategy is the index delta (for a description of index delta, see Chapter 3, "Risk Management"). When setting values for different parameters, the developer should watch that the index delta of the portfolio is equal or close to zero. Other risk measures, including Value at Risk (VaR), asymmetry coefficient, and loss probability, are used as additional risk management tools (these gauges will be discussed in detail in Chapter 3).

1.2.2. Points and Boundaries of Delta-Neutrality

In the preceding section we described the main elements of delta-neutral strategies. One can easily see that even the simple basic version of this strategy has a fairly large number of parameters for which definite values have to be set and fixed. The presence of just several parameters means that there is a huge number of different variants to combine their values (it increases according to the power law). The situation is further complicated by the fact that *for most combinations of parameter values, delta-neutrality is not attainable.*

In the basic delta-neutral strategy there are three parameters that directly influence the structure and the main properties of the portfolio:

- The threshold value of the criterion that is used to generate signals for positions opening
- The range of strike prices that are used to create option combinations
- Option series used to create combinations (this parameter determines the time left to the option expiration date)

When setting these parameters, the developer of the trading strategy must consider their influence on such important portfolio characteristics as the relative frequency of long and short positions, diversification, and risk indicators. However, the relationships

between the portfolio delta and the value of each of these three parameters (and their various combinations) must be established in the first place. After all, if for most acceptable parameter values the portfolio index delta deviates from zero significantly, the delta-neutral strategy can hardly be created.

Each combination of parameter values for which the delta-neutrality condition is fulfilled (the portfolio delta is zero) will be called the point of delta-neutrality. The whole set of such points composes the boundary of delta-neutrality.

Let us consider several examples of determining the points of delta-neutrality. Suppose that in order to generate trading signals we evaluate the set of option combinations created for all stocks of the S&P 500 index. The valuation is performed using the "expected profit on the basis of lognormal distribution" criterion (according to the procedure described in the preceding section). Assume that the "strikes range" parameter is fixed at 10% (option combinations are constructed using strikes situated in the range of 10% of the underlying asset price). Several values of the "time left to options expiration" parameter are examined: one week and one, two, and three months until expiration. To determine the points of delta-neutrality on January 11, 2010, we generated trading signals for the following expiration dates: January 15, 2010 (one week to expiration), February 19, 2010 (one month to expiration), March 19, 2010 (two months to expiration), and April 16, 2010 (three months to expiration).

The points of delta-neutrality have to be determined for the whole range of criterion threshold values. To perform this procedure, we need to establish the relationship between the index delta and the criterion threshold. Figure 1.2.1 shows these relationships for four expiration dates. Points that lie at the intersection of the chart line and the horizontal axis are delta-neutral (in these points the index delta equals zero). Accordingly, each intersection point shows the threshold value for which the condition of delta-neutrality is met (the criterion threshold is equal to the coordinate at the horizontal axis).

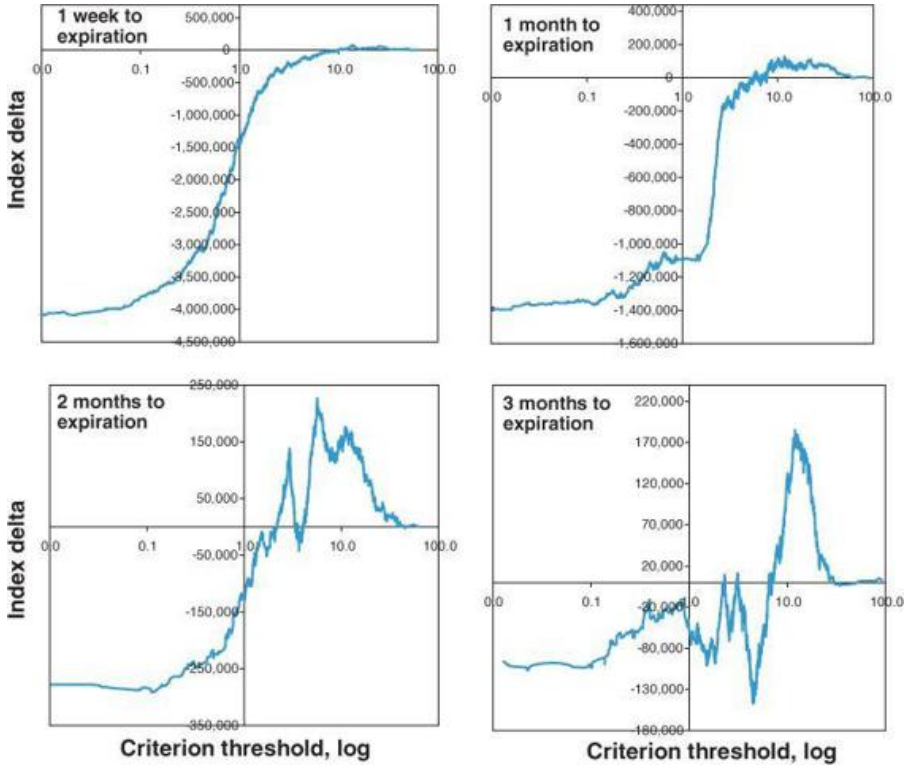

Figure 1.2.1. Relationships between the index delta and the criterion threshold parameter. The "strikes range" parameter is fixed at 10%. Each chart shows the relationship for one of four values of the "time to expiration" parameter. Each intersection of the chart line and the horizontal axis represents the point of delta-neutrality.

In four particular cases shown in Figure 1.2.1, delta-neutrality is achieved with the threshold values ranging from 2% to 10% (criterion values are expressed as a percentage of the investment amount). In the case when only one week is left until expiration (the upper-left chart of Figure 1.2.1), there is only one delta-neutrality point (situated at the threshold value equal to 9%). This means that if we build option combinations using this time series and using strike prices located inside the range {underlying asset price $\pm 10\%$}, and if we select combinations for which the criterion value is equal to or greater than 9%, the portfolio will be delta-neutral.

A great number of delta-neutrality points exist when the portfolio is created using options with one month left to expiration (since the delta line crosses the horizontal axis many times). Intersections are located in a rather narrow range of criterion threshold values. Figure 1.2.2 shows this range on a larger scale, which enables us to

distinguish each separate point of delta-neutrality. Altogether there are 16 such points situated within the interval 5% to 8%. (In other words, the criterion threshold values for which delta-neutrality condition is fulfilled are located inside the range of 5% to 8%.) In the case when the value of the "time left to options expiration" parameter was set at two months, there were three points of delta-neutrality (the bottom-left chart in Figure 1.2.1). When it was set at three months, five delta-neutrality points were observed (the bottom-right chart in Figure 1.2.1).

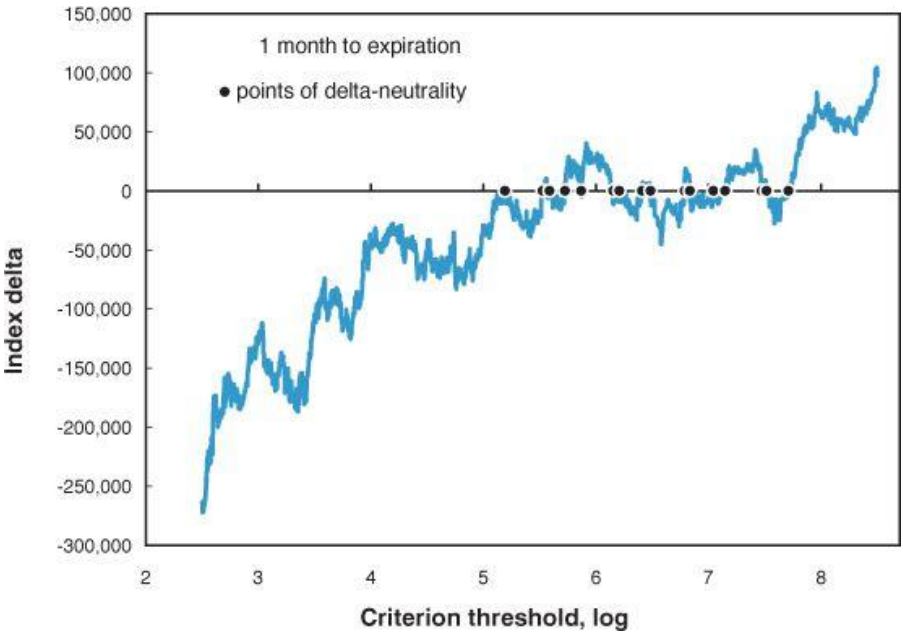

Figure 1.2.2. Relationship between the index delta and the criterion threshold parameter when the portfolio is created using options with one month left until expiration. Only the range of criterion threshold values with points of delta-neutrality is shown.

It is important to note that as the value of the "criterion threshold" parameter increases, the index delta of the portfolio consisting of options with close expiration dates changes in a very wide range. At the same time, the delta of the portfolio consisting of distant option series changes in a much narrower interval (compare the top-left and bottom-right charts in Figure 1.2.1). The reason is that, all other things being equal, the option delta increases as the expiration date approaches (if the option is in-the-money and there is little time left to expiration, its delta verges toward +1 for call options or –1 for put options). This brings us to an important conclusion: *When a portfolio is created using options with a close expiration date, a small deviation from a*

given combination of parameter values (for which a portfolio is expected to be delta-neutral) brings about greater deviation from delta-neutrality than when long-term options are used.

Now we turn to the procedure of finding the boundaries of delta-neutrality. First we fix one parameter, time to expiration, and examine all possible combinations of values for two other parameters, criterion threshold and range of strike prices. To perform this procedure, we need to calculate the index delta for all {criterion threshold × strikes range} variants at the whole range of their acceptable values. Then this data should be presented in the form of a topographic map where horizontal and vertical axes correspond to the values of parameters under investigation and each point on the map expresses the height mark corresponding to the index delta value. Points with the same height marks on this map represent isolines. *The isoline going through zero level is a delta-neutrality boundary.*

Figure 1.2.3 shows an example of such a topographic map. In this example the "time to expiration" parameter is set at one week. (We used the same data as for the top-left chart in Figure 1.2.1; the portfolio creation date is January 11, 2010, and the expiration date is January 15, 2010.) The delta-neutrality boundary traverses the map diagonally from the top-left corner to the bottom-right corner. Portfolios with positive values of index delta are situated to the right of the boundary; portfolios with negative delta are located to its left. When the criterion threshold value is very low (left edge of the map), the portfolios delta reaches highly negative values.

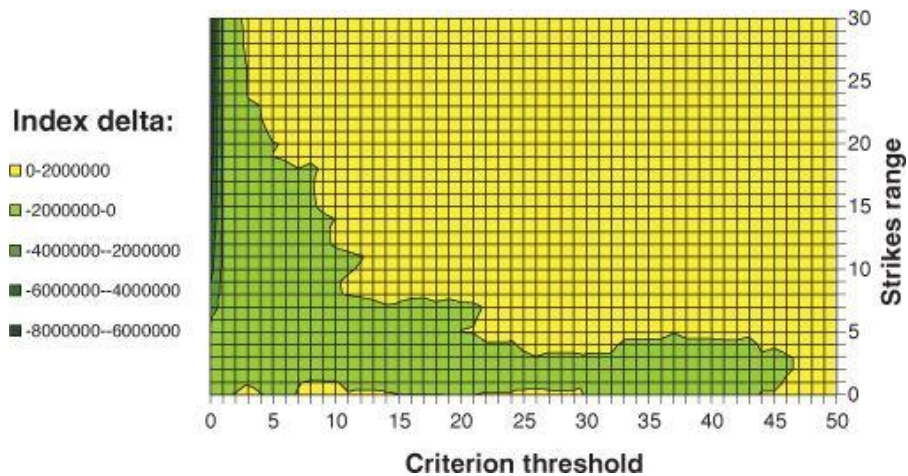

Figure 1.2.3. An example of the topographic map showing the boundary of delta-neutrality.

Note that the top-left chart in Figure 1.2.1 corresponds to the horizontal line on the map (in Figure 1.2.3) passing through the

strikes range of 10%. That is, if we make an incision at this line and depict its side view, we will obtain a profile matching the index delta line shown in Figure 1.2.1. Thus, *the whole set of delta-neutrality points can be consolidated to form the boundaries of delta-neutrality.*

It follows from the topographic map presented in Figure 1.2.3 that the delta-neutral portfolio can be created with a rather large number of {criterion threshold × strikes range} alternatives. For example, we can build the portfolio using option combinations for which the criterion value is higher than 5%, and the range of strike prices is fairly wide (price ± 20%). At the same time, we may prefer another delta-neutral portfolio consisting of combinations with high criterion values (e.g., higher than 15%). In this case, however, we must limit ourselves to a rather narrow range of strike prices (price ± 6%).

This method of finding delta-neutrality boundaries is based on the visual analysis. It was selected to make the description simple and clear. In practice, however, we do not need to visualize boundaries; they can be determined analytically using computer algorithms. At the same time, building such topographic maps may be useful in perceiving parameters' interrelationships and selecting their acceptable values ranges.

Undoubtedly, neither trading strategy can be grounded on a single instance analysis similar to the one depicted in Figure 1.2.3. The main quality of automated trading is that all decision-making algorithms are based on broad statistical material. It should be considered whether the location of delta-neutrality boundaries depends on market conditions. It would be logical to suppose that in periods of high volatility the boundaries are different from those in periods of a calm market. For that very reason we now pass on to analyzing delta-neutrality boundaries on the basis of broader statistical data, including calm as well as extreme periods.

1.2.3. Analysis of Delta-Neutrality Boundaries

In this section we analyze the effect of time left to options expiration and that of market conditions prevailing at the moment of portfolio creation on the delta-neutrality boundaries. Two time periods and two states of underlying assets market will be studied. Time intervals will be fixed at one week and two months between the position opening and the expiration date. For each of these time intervals, delta-neutrality boundaries will be examined during a calm market with low volatility as well as during extremely volatile periods. For each of these four {time period × state of market} variants, we will build 12 delta-neutrality boundaries (relating to different expiration dates). For the calm market we will use data from

March 2009 to February 2010. For the volatile market the data referring to the last financial crisis (January 2008 to December 2008) will be used.

The top chart in Figure 1.2.4 shows 12 delta-neutrality boundaries for portfolios created during the calm period using options with the nearest expiration date. Although boundaries do not coincide (which is quite natural since they all correspond to different expiration months), the general pattern is clear enough and allows making several important conclusions. Firstly, all boundaries are situated approximately in the same area. Secondly, most boundaries have a similar shape, which is more or less stretched toward high criterion threshold values. Thirdly, most boundaries are located in the area of the rather narrow strikes range (although in some cases they have "appendices" extending toward wider ranges). Such a form of boundaries indicates that *in periods of calm market, the fulfillment of a delta-neutrality condition for portfolios consisting of options with a close expiration date is possible in quite a wide range of criterion threshold values, but this requires using a narrow range of strike price.*

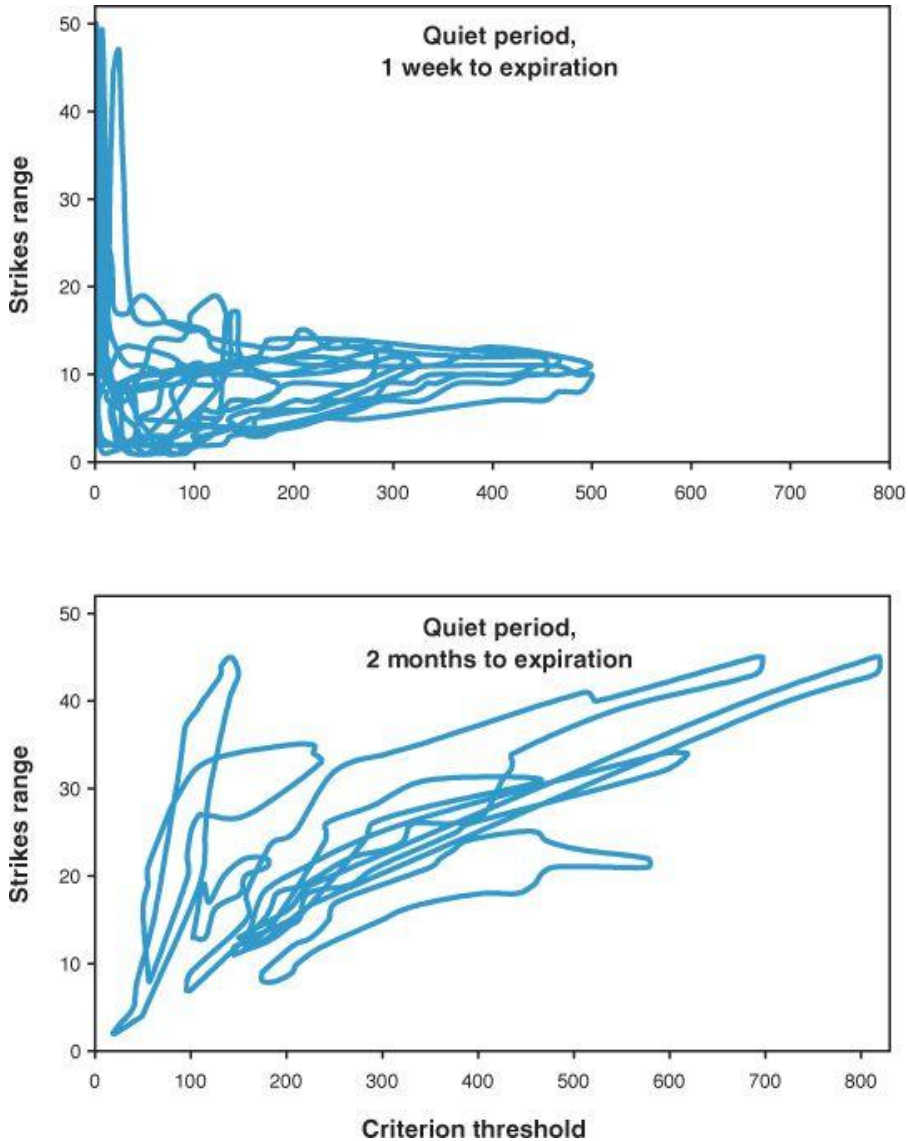

Figure 1.2.4. Delta-neutrality boundaries of portfolios created during a calm market period. The top chart shows portfolios constructed of options with one week left to expiration. The bottom chart shows portfolios created using options with two months to expiration. Each boundary refers to a separate expiration date.

Delta-neutrality boundaries relating to portfolios created in a calm period using options with a relatively distant expiration date have a fundamentally different appearance (the bottom chart in Figure 1.2.4). First, it should be noted that delta-neutrality boundaries were obtained in only 8 out of 12 cases. In the other 4 cases delta-neutrality

appeared to be unattainable at any of the points. Second, boundaries related to different expiration months overlap to a smaller extent as compared to the boundaries observed for portfolios consisting of options with the nearest expiration date (compare the bottom and top charts in Figure 1.2.4). This means that *as the time period until options expiration increases, the variability of delta-neutrality boundaries also increases.* Nonetheless, it is possible to distinguish an area where most of the boundaries are situated. It represents a rather wide band passing diagonally from the low criterion threshold value and narrow strikes range to the high threshold value and wide strikes range. This pattern differs widely from the case of portfolios constructed of options with the nearest expiration date where the delta-neutrality area was nearly parallel to the criterion threshold axis (the top chart in Figure 1.2.4). These properties of delta-neutrality boundaries suggest that *the selection of a higher criterion threshold for creation of a delta-neutral portfolio (in calm markets and using options with distant expiration dates) requires widening the strikes range.*

Now we turn to the period of high volatility. When the portfolio consisted of options with the nearest expiration date, delta-neutrality was achieved in 7 out of 12 cases (the top chart in Figure 1.2.5). For portfolios constructed using long-term options, the delta-neutrality was even less attainable (in only 5 out of 12 cases; see the bottom chart in Figure 1.2.5). In one of these five cases, the boundary is very short. During high-volatility periods, the shape of delta-neutrality boundaries depends on the time until expiration to a lesser extent than during calm markets. This conclusion follows from the similarity of the top and bottom charts in Figure 1.2.5 (the charts in Figure 1.2.4 differ to a much greater extent). The only difference between the time-to-expiration periods is higher positioning of boundaries along the vertical axis observed for portfolios with a more distant expiration date (the bottom chart in Figure 1.2.5). This indicates that *using a wider strikes range is required to achieve delta-neutrality during volatile markets if portfolios are to be constructed of more distant options.*

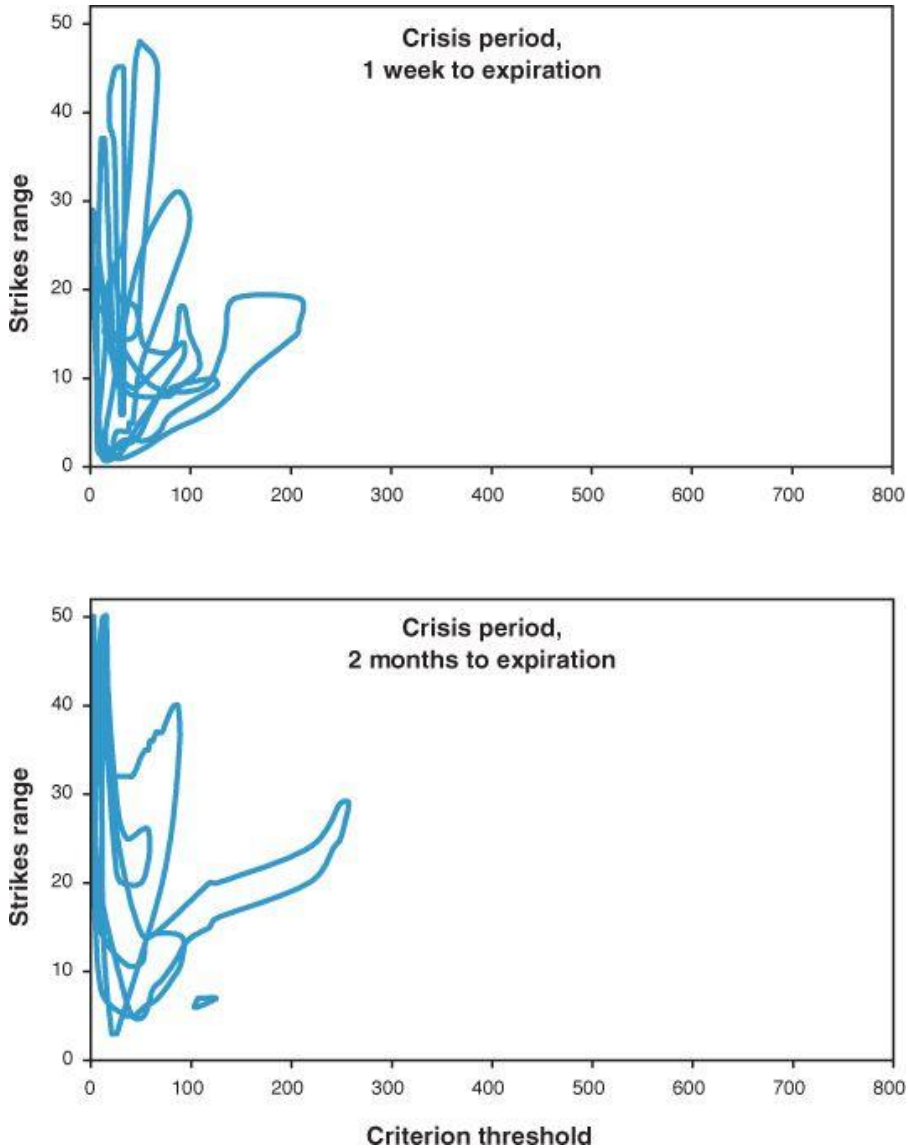

Figure 1.2.5. Delta-neutrality boundaries of portfolios created during a volatile market period. The top chart shows portfolios constructed of options with one week left to expiration. The bottom chart shows portfolios created using options with two months to expiration. Each boundary corresponds to a separate expiration date.

During extreme market fluctuations, the shape of delta-neutrality boundaries looks similar to the shape of boundaries obtained in the period of calm markets. However, these boundaries do not stretch that far along the criterion threshold axis (compare Figures 1.2.4 and 1.2.5). This means that *during volatile market periods, building delta-*

neutral portfolios using option combinations with exclusively high criterion values is impossible.

1.2.4. Quantitative Characteristics of Delta-Neutrality Boundaries

Visual analysis of boundaries similar to those shown in Figures 1.2.4 and 1.2.5 is vivid, but is inevitably prone to the influence of subjective factors depending on the individual perception of the researcher. Besides, visual analysis is limited by human abilities and cannot cover big datasets. To create an automated trading system, the developer should be able to analyze delta-neutrality boundaries on the basis of quantifiable numerical characteristics.

We developed a methodology for describing delta-neutral boundaries by means of four quantitative characteristics:

1. *The criterion threshold index* determines the position of the delta-neutrality boundary relative to the criterion threshold axis. This characteristic is calculated by averaging the coordinates on the horizontal axis of the topographic map for all points of the delta-neutrality boundary.

2. *The strikes range index* determines the position of the delta-neutrality boundary relative to the strikes range axis. This indicator is calculated by averaging the coordinates on the vertical axis of the topographic map for all points of the delta-neutrality boundary.

3. *The length of the delta-neutrality boundary* describes the length of the boundary. The value of this indicator is equal to the number of points composing the delta-neutrality boundary.

4. *The delta-neutrality attainability* characterizes the possibility of creating delta-neutral portfolios. This indicator expresses the percentage of cases for which delta-neutrality is achievable.

Let us apply this methodology to the delta-neutrality boundaries shown in Figures 1.2.4 and 1.2.5. Criterion threshold and strikes range indexes are shown in Figure 1.2.6. Here, each boundary that was presented earlier as isoline (see Figures 1.2.4 and 1.2.5) is reduced to a single point. Each of these points can be viewed as a kind of a center of the area bounded by the delta-neutrality boundary. Of course, reducing the boundary to a single point cuts the quantity of information contained in the original data. Nonetheless, almost all conclusions made earlier on the basis of visual analysis of Figures 1.2.4 and 1.2.5 can equally be drawn from the data presented in Figure 1.2.6.

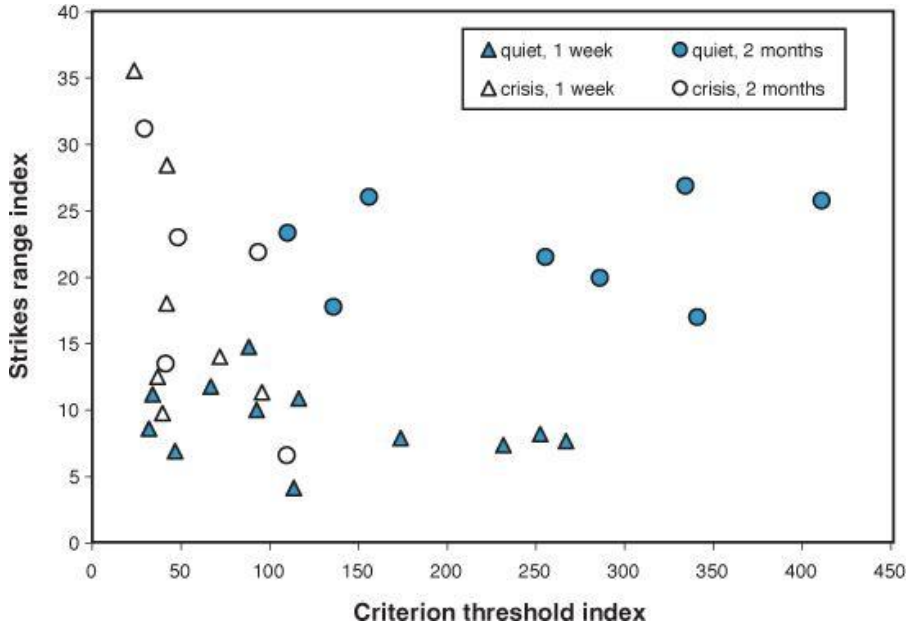

Figure 1.2.6. Presentation of delta-neutrality boundaries as single points with coordinates corresponding to criterion threshold and strikes range indexes. Each point corresponds to one of the boundaries shown in Figures 1.2.4 and 1.2.5.

Reducing delta-neutrality boundaries to points enables simultaneous comparison of a large number of boundaries within the scope of one complex analysis. This can be illustrated by an example of expanded analysis covering ten years of historic data. Two samples were taken from the database containing the prices of options and their underlying assets from March 2000 to April 2010. One sample corresponds to the period of low volatility (when historic volatility did not exceed 15%); the other one, to the period of high volatility (with historic volatility of more than 40%). Within each sample, we created two portfolio variants. One of the variants consisted of the nearest options expiring in one to two weeks; another variant consisted of distant options expiring in two to three months. Both portfolio versions were created using signals for positions opening generated by the basic delta-neutral strategy.

Figure 1.2.7 shows criterion threshold and strikes range indexes for each of the portfolios obtained in this study. Analysis of these data (here we rely on visual analysis, though in practice it can be replaced by a computer algorithm) allows us to draw a number of important conclusions regarding the relationships among delta-neutrality boundaries, market volatility, and time left until options expiration:

• *During calm periods, delta-neutrality boundaries are situated in*

the area of low criterion threshold values (regardless of the time period left until expiration). This conclusion is based on the location of filled circles and triangles in the left part of Figure 1.2.7.

• *During high-volatility periods, the position of delta-neutrality boundaries is quite unsteady in relation to the criterion threshold axis,* but in most cases has a propensity for the area of low values. This conclusion is based on the location of contour circles and triangles along the horizontal axis (with their predominance at the left part of this axis) in Figure 1.2.7.

• *The time left until options expiration does not influence the position of delta-neutrality boundaries relative to the criterion threshold axis.* Delta-neutrality can be achieved in the widest interval of threshold values. This conclusion is based on the location of circles and triangles along the horizontal axis in Figure 1.2.7.

• *For portfolios consisting of options with the nearest expiration date, delta-neutrality is attainable on the condition that a narrow strikes range is used.* This conclusion (which is accurate for both calm and volatile markets) is based on the location of filled and contour triangles in the area of low values relative to the vertical axis in Figure 1.2.7.

• *If portfolios are created during volatile periods using distant options, delta-neutrality boundaries are situated mostly in the area of average strikes range values.* This conclusion is based on the location of most of the contour circles approximately in the middle of the vertical axis in Figure 1.2.7.

• *During low-volatility periods, only portfolios (consisting of distant options) that were created using a wide strikes range can be delta-neutral.* This conclusion is based on the location of the filled circles in the area of high values of the vertical axis in Figure 1.2.7.

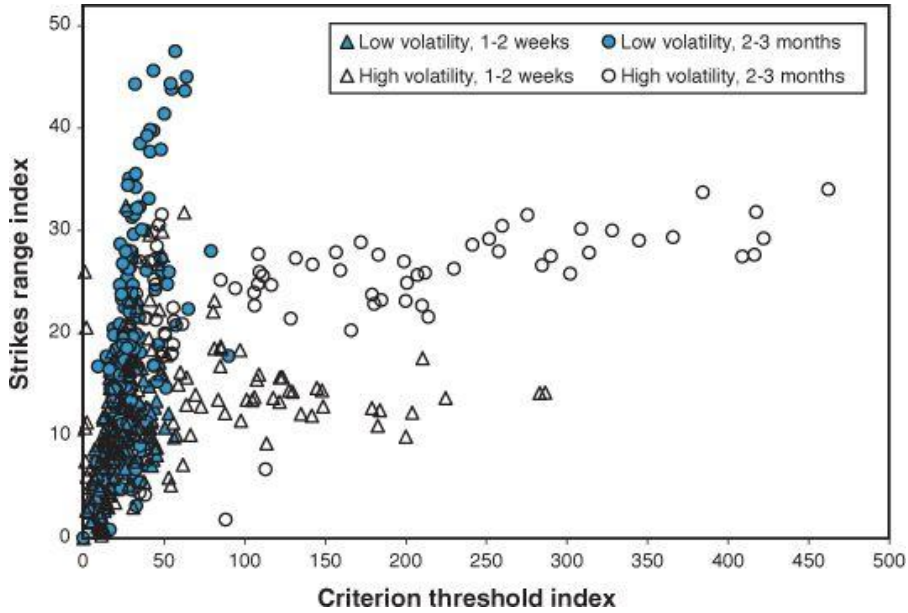

Figure 1.2.7. Presentation of delta-neutrality boundaries as single points with coordinates corresponding to the criterion threshold and strikes range indexes. The data relates to a ten-year historical period.

Now we turn to the length of delta-neutrality boundaries. As mentioned earlier, this characteristic expresses the number of points composing the boundary. The longer the boundary, the more variants of delta-neutral portfolio can be created by manipulating criterion threshold and strikes range values.

First we analyze the interrelationships between the boundary length and the two characteristics that were discussed previously: criterion threshold and strikes range indexes. Figure 1.2.8 shows the relationship between the boundary length and its position in the topographic map. The longest boundaries are situated in the area of high criterion threshold values and low values of strikes range. This means that *in order to obtain a longer delta-neutrality boundary, the strategy should use a narrow strikes range and select combinations with high criterion values.* This will broaden the choice of delta-neutral portfolios, which, in turn, will allow selecting a portfolio with characteristics thoroughly satisfying the requirements of the trading system developer.

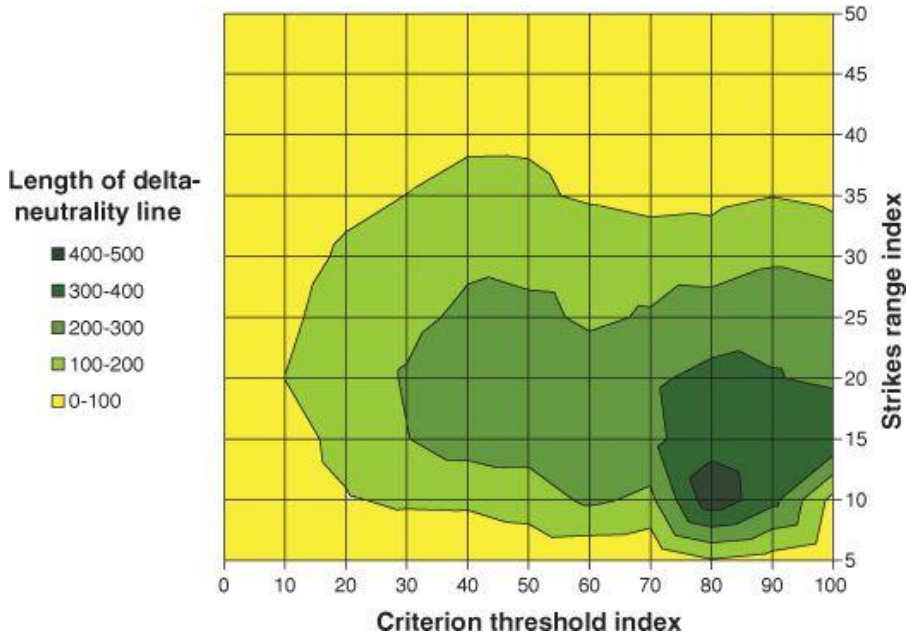

Figure 1.2.8. Relationships between the length of the delta-neutrality boundary and the indexes of criterion threshold and strikes range. The data relates to a ten-year historical period.

The dependence of the delta-neutrality boundaries length on market volatility and time left until options expiration is presented in Figure 1.2.9. The longest boundaries are observed when the volatility is high and the time left to options expiration is rather short (about 20 days). If the time to expiration is less than 20 days, the boundary length decreases. More distant options (more than 20 days to expiration) also shorten the delta-neutrality boundary. Thus, *during high-volatility periods, the largest number of variants of delta-neutral portfolios exists when the time of positions opening is separated from the options expiration date by about 10 to 30 trading days.*

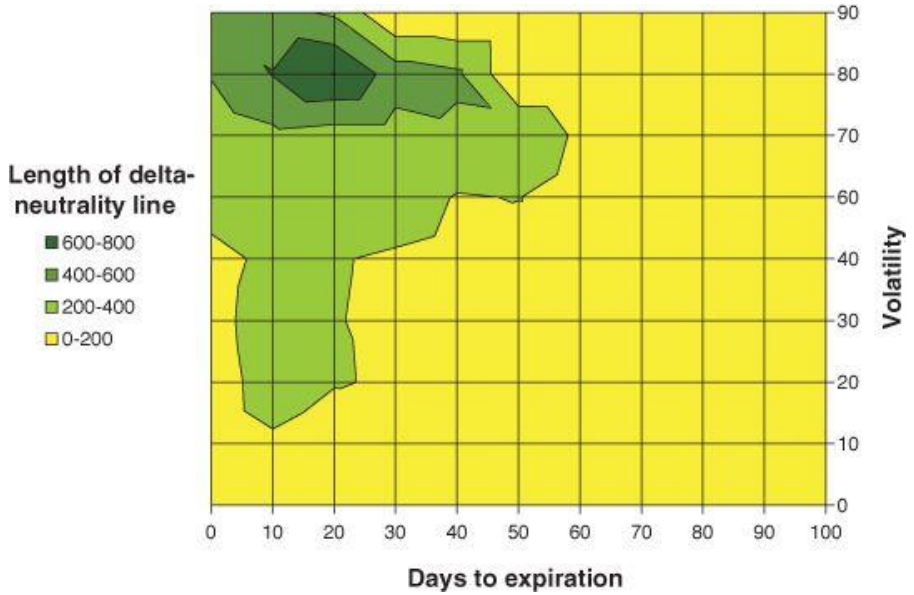

Figure 1.2.9. Relationships between the length of the delta-neutrality boundary, the number of days left until options expiration, and market volatility. The data relates to a ten-year historical period.

As volatility decreases, the length of the delta-neutrality boundary also decreases (see Figure 1.2.9). This relationship does not depend on the number of days left to options expiration. However, when positions are opened shortly before the expiration, the boundary shortens more slowly. Even when historic volatility decreases to about 20%, the length of the delta-neutrality boundary still remains within average values. Under extremely low volatility the boundary shortens to a minimal length at the whole range of time-to-expiration values. In some cases the shortening of the delta-neutrality boundaries can be so drastic that the whole boundary shrinks to a tiny group of points. An example of such extremely short boundary was shown earlier for the portfolio created during the crisis period using long-term options (the bottom chart in Figure 1.2.5).

In extreme cases delta-neutrality boundaries may shrink to zero. In such cases we say that delta-neutrality is unattainable. Let us examine whether the attainability of delta-neutrality depends on the number of days left until options expiration and on market volatility. To test this, we again took two samples from the ten-year database. One of them was taken from the data pertaining to a calm market period (historic volatility $< 15\%$); another one, from the high-volatility data (historic volatility $> 40\%$). Two portfolio variants were created within each sample: one consisting of short-term options and another one

constructed using distant options (one to two weeks and two to three months to expiration, respectively). According to the definition given earlier, delta-neutrality attainability was estimated through the number of cases (portfolios) when delta-neutrality was achieved at least at one point. It was expressed as a percentage of the total number of cases.

Statistically significant inverse relationships were discovered between delta-neutrality attainability and the number of days left until options expiration (for a calm period $t = 9.12$, $p < 0.0001$; for a volatile period $t = 15.24$, $p < 0.0001$; see Figure 1.2.10). Delta-neutrality appeared to be attainable in almost all cases when portfolios were created just several days before expiration. Moving away from expiration reduces delta-neutrality attainability. When portfolios were created using long-term options (with more than 100 working days left until expiration), delta-neutrality was attainable in only 40% to 80% of cases (depending on market volatility). Values of determination coefficients (0.60 for a high-volatility period and 0.35 for a calm period) indicate that 35% to 60% of variability in delta-neutrality attainability is explained by the time period from positions opening to options expiration. Thus, *using the nearest options guarantees that most portfolios will be delta-neutral.*

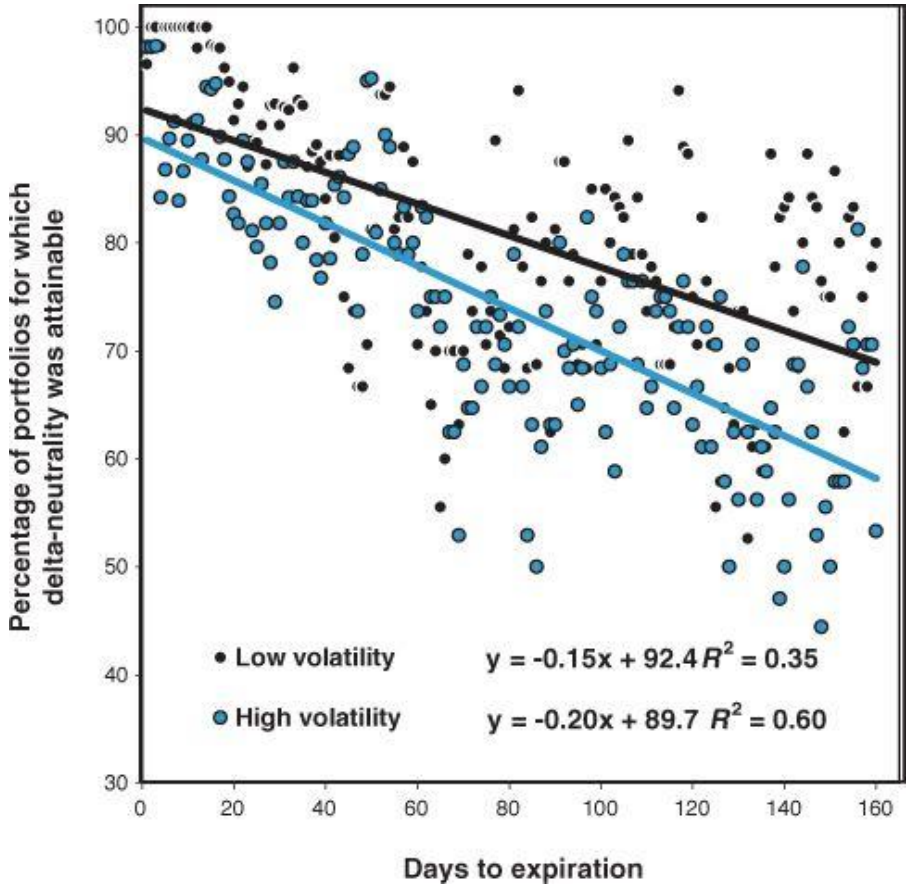

Figure 1.2.10. Relationship between delta-neutrality attainability and the number of days left until options expiration during calm and volatile markets. The data relates to a ten-year historical period.

Market volatility influences the relationship between delta-neutrality attainability and the number of days left to options expiration. This is evident from the distribution of data points in the two-dimensional coordinate system and the divergence of regression lines relating to different volatility levels (see Figure 1.2.10). Slope coefficient reflecting the relationship that existed during a volatile market is less than the coefficient relevant to a calm market. This difference in regression line slopes is statistically significant ($t = 2.76$, $p < 0.0065$). Hence, *delta-neutrality attainability depends on the state of the market:* under high-volatility conditions the probability to attain delta-neutrality is lower. The impact of volatility on the attainability of delta-neutrality increases as the time left until options expiration increases.

Table 1.2.1 summarizes the analysis of delta-neutrality

characteristics conducted in this section. It presents the effects of market volatility and time period from portfolio creation to options expiration on the delta-neutrality border location, its length, and delta-neutrality attainability. This table can be used in the following way. Let us presume that current market volatility is high and, for some reason or other, the portfolio is created using options with close expiration date. In such circumstances, we can deduce the following:

- Delta-neutrality can be achieved within the broad interval of criterion threshold values (denoted in the table as "variable criterion threshold").
- To obtain a delta-neutral portfolio, we have to use strike prices situated close to the current underlying asset price (denoted in the table as "narrow strikes range").
- The probability to attain delta-neutrality in the given circumstances is high (denoted in the table as "high delta-neutrality attainability").
- There is a great number of variants of delta-neutral portfolio (denoted in the table as "long delta-neutrality line").

Table 1.2.1. Dependence of delta-neutrality characteristics (delta-neutrality boundary [DNB] location, DNB length, and delta-neutrality attainability [DNA]) on market volatility and time from portfolio creation to options expiration.

		Volatility			
		High		Low	
Time to Expiration	Short	DNB location: DNB location: DNB length: DNA:	*variable criterion threshold* *narrow strikes range* *long* *high*	DNB location: DNB location: DNB length: DNA:	*low criterion threshold* *narrow strikes range* *medium* *high*
	Long	DNB location: DNB location: DNB length: DNA:	*variable criterion threshold* *medium strikes range* *short* *low*	DNB location: DNB location: DNB length: DNA:	*low criterion threshold* *wide strikes range* *short* *low*

1.2.5. Analysis of the Portfolio Structure

In previous sections we discussed the issue of delta-neutrality attainability and analyzed various factors affecting the position and length of delta-neutrality boundaries. We determined how to increase the number of available variants of the delta-neutral portfolio by manipulating three main strategy parameters. All effects were examined for both volatile and calm markets.

Often we have a choice between (at least) several alternative

combinations of parameter values that can be equally used to create a delta-neutral portfolio. However, the structure and properties of these portfolios will be different. This diversity invites the question of how to select such a combination of parameters that would produce a portfolio satisfying the requirements of the strategy developer as much as possible.

To describe the general structure and the main properties of an option portfolio, we will use the following characteristics:

- The number of combinations in a portfolio (which expresses the diversification level, the number of trades, and, hence, the amount of slippage and operational costs)
- The number of different underlying assets in a portfolio (which represents an additional indicator of the diversification level and the number of trades)
- Proportions of long and short combinations in a portfolio (which describe portfolio structure)
- Proportions of straddles and strangles in a portfolio (which describe portfolio structure)
- The extent of portfolio asymmetry (in addition to index delta, this characteristic expresses, in a different way, the degree of market-neutrality of the portfolio)
- Loss probability and VaR (which express portfolio risk)

In this section we analyze the effects of four main parameters (strikes range, criterion threshold, time to expiration, and market volatility) on these characteristics. Since, as we will show shortly, many characteristics change significantly depending on the criterion threshold and strikes range values, the analysis will be limited to the interval of values of these parameters from 0% to 25%. Otherwise, it would not be possible to distinguish the tendencies in changes of the characteristics in the charts.

This study is based on data pertaining to both a calm period with low volatility and a crisis period with extreme market fluctuations. Two time intervals from the moment of portfolio creation to options expiration are analyzed (one week and two months). Portfolios related to a calm market were created January 11, 2010; the expiration dates are January 15, 2010 (one-week options), and March 19, 2010 (two-month options). Portfolios related to volatile market were created November 17, 2008; expiration dates are November 21, 2008 (one-week options), and January 16, 2009 (two-month options).

Number of Combinations in the Portfolio

The shapes of surfaces shown in Figure 1.2.11 demonstrate the extent of influence exerted by different parameters on the characteristic under scrutiny. The number of combinations in the portfolio reaches its maximum when the limitations imposed on strategy parameters are minimal: The criterion threshold is the lowest (this means that all combinations with positive expected profit are included in the portfolio); the strikes range is wide (meaning that the maximum number of strike prices is used to create combinations); and the time left until the expiration date is long. This holds true for both calm and volatile markets.

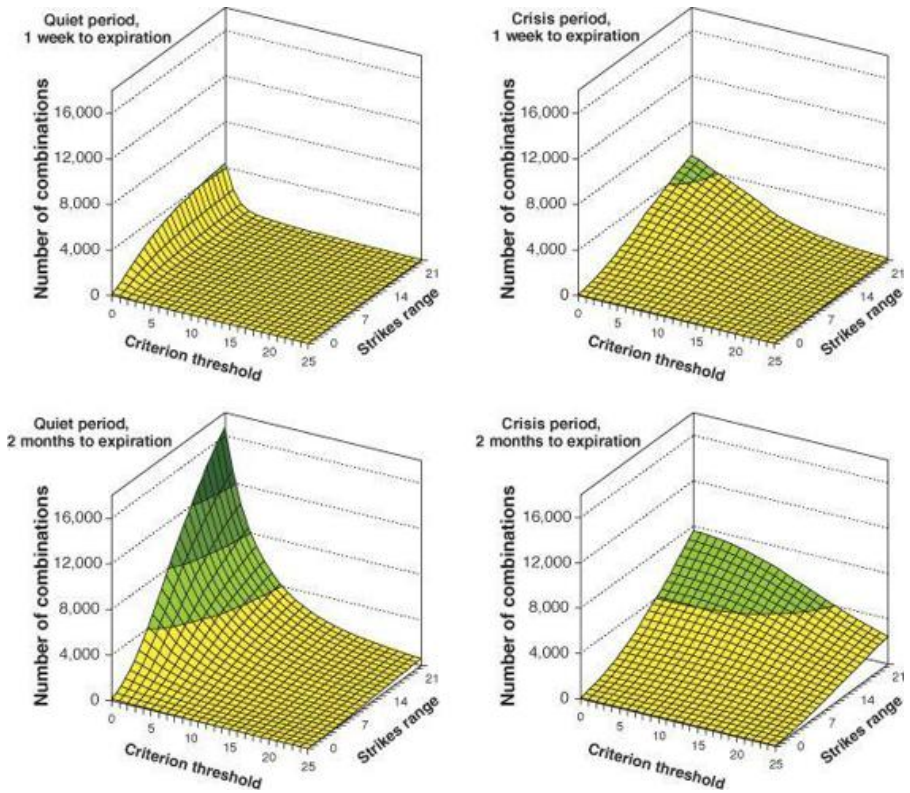

Figure 1.2.11. Dependence of the number of combinations included in the portfolio on the criterion threshold and strikes range. Two market conditions (calm and crisis) and two time periods from the time of portfolio creation until options expiration (one-week and two-month) are shown.

Charts relating to the calm market (the left-hand charts in Figure 1.2.11) show that increasing the criterion threshold leads to an exponential decrease in the number of combinations in the portfolio. Narrowing the strikes range also brings about a decrease in the quantity of combinations, though this decrease is more gradual. The surface relating to portfolios created using the nearest options during

the volatile market (the top-right chart in Figure 1.2.11) shows the approximately equal effect of both parameters on portfolios diversification. When long-term options are used during crisis periods (the bottom-right chart in Figure 1.2.11), a rising criterion threshold leads to a smoother decrease in the number of combinations than the narrowing of strikes range does.

In periods of calm market, the least number of combinations is obtained for portfolios created close to the expiration date (the top-left chart in Figure 1.2.11). When more distant options are used (the bottom-left chart in Figure 1.2.11), the number of combinations increases significantly. The same trends were observed during volatile periods. However, the difference between portfolios created close to expiration (the top-right chart in Figure 1.2.11) and far from expiration (the bottom-right chart in Figure 1.2.11) is not as substantial as was noted in the case of the calm market. Besides, portfolios constructed of the nearest options contain fewer combinations during the calm market than during the crisis period (the top charts in Figure 1.2.11). The opposite is true for portfolios created using distant options: In calm market conditions they consist of more combinations than during volatile periods (the bottom charts in Figure 1.2.11). These patterns hold for all criterion thresholds and all strikes range values, though they are more pronounced at lower criterion thresholds and wider strike ranges.

Number of Underlying Assets in the Portfolio

This characteristic represents another important indicator of diversification. Since fluctuations of underlying asset prices are among the main factors influencing profits and losses of delta-neutral option portfolios, diversification can reduce unsystematic risk significantly. At the same time, redundant diversification may have a negative impact since it requires a large number of excessive trades (thereby increasing losses due to slippage and operational costs) while bringing about only a small risk decrease.

During calm periods, the number of underlying assets in a portfolio depends only on the criterion threshold and the time left to options expiration. The breadth of strikes range does not influence diversification of the portfolio consisting of the nearest options (the top-left chart in Figure 1.2.12). When distant options were used to construct a portfolio, only a weak effect of strikes range can be distinguished in the shape of the surface (the bottom-left chart in Figure 1.2.12). Under the lowest criterion threshold value, the number of underlying assets may be very high, if a portfolio is created using options with the nearest expiration date. For the criterion threshold value of 1%, this number reaches about 400 out of a possible 500

(remember that this study is limited to stocks included in the S&P 500 index). However, even a slight increase in the criterion threshold leads to a dramatic nonlinear decrease in the quantity of underlying assets. When the threshold value is raised to 3%, the number of underlying assets falls to 50; for an 8% threshold it does not exceed 20. A similar pattern is observed when the portfolio was created using distant options. However, in this case a bit more underlying assets correspond to each {criterion threshold × strikes range} combination. This conclusion is supported by the less concave surface shape of the bottom-left chart in Figure 1.2.12. It should also be noted that when two-month options are used to construct the portfolio during a calm market, the widening of the strikes range entails the increase (although insignificant) in the quantity of underlying assets.

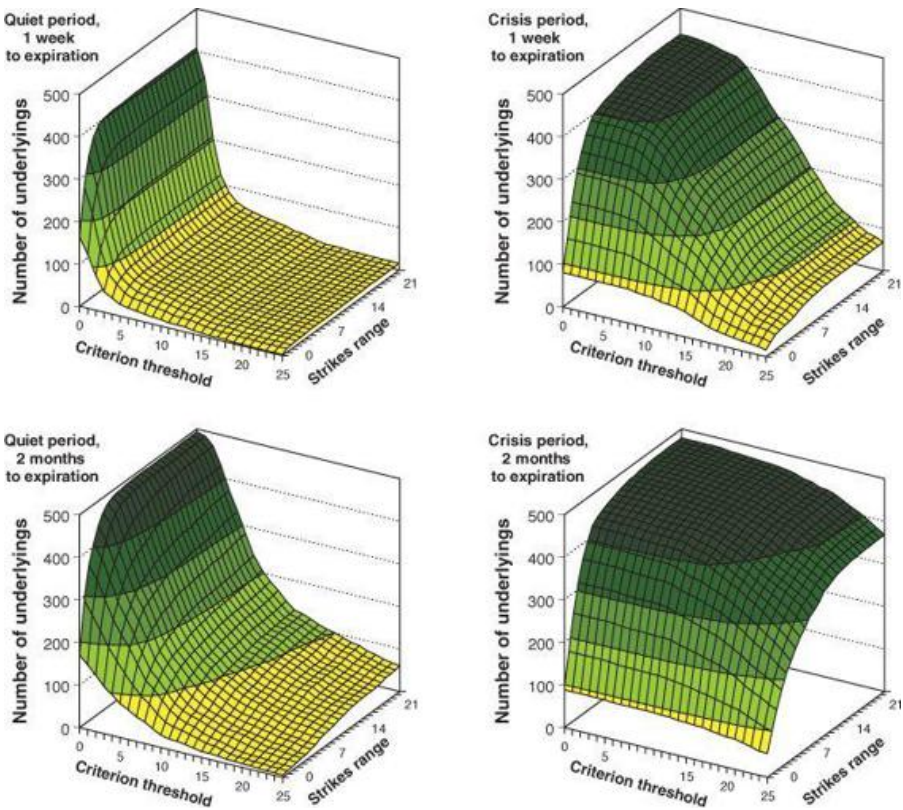

Figure 1.2.12. Dependence of the number of underlying assets on the criterion threshold and strikes range. Two market conditions (calm and crisis) and two time periods from the time of portfolio creation until options expiration (one-week and two-month) are shown.

In periods of high volatility, the picture is quite different. When the portfolio consists of the nearest options, the criterion threshold and

strikes range influence the number of underlying assets similarly. Diversification reaches its maximum at lower criterion threshold values combined with a wider strikes range. The flat plateau on the top-right chart in Figure 1.2.12 attests to this conclusion. Increasing the criterion threshold and curtailing the strikes range lead to a sharp decrease in the quantity of underlying assets. When combinations are constructed of long-term options, the effect of the criterion threshold on portfolio diversification is insignificant. In such circumstances the main parameter that determines the number of underlying assets is the strikes range (the bottom-right chart in Figure 1.2.12).

Proportions of Long and Short Combinations

This characteristic expresses one of the fundamental qualities of the options portfolio. On the one hand, the proportions of long and short positions directly influence the risk as well as the potential profitability of the portfolio. (Recall that profitability of short combinations is limited while their loss is potentially boundless. On the contrary, long combinations have limited downside risk and unlimited upside potential.) On the other hand, the range of acceptable values for this characteristic is determined not only by its influence on prospective risk and return, but also by many additional factors. They include various restrictions imposed by in-house investment policies and external regulators (which may be considered as objective limiting factors), as well as psychological constraints of the strategy developer (subjective limiting factor).

During high-volatility periods, all portfolios are composed mostly of short combinations (apart from rare exceptions). For this reason we do not present the charts related to crisis periods in Figure 1.2.13. The prevalence of short combinations during extreme market fluctuations is a common occurrence. This happens mostly due to the fact that option premiums (implied volatility) rise rapidly during crises, while historical volatility (which is estimated using historical price series that includes both current [volatile] and previous [calm] subperiods) increases more slowly. Consequently, most criteria have higher values for short combinations. Nevertheless, even during high-volatility periods, delta-neutral portfolios may include a certain (though limited) number of long combinations. (After all, the examples presented in this section are limited to only two expiration dates.)

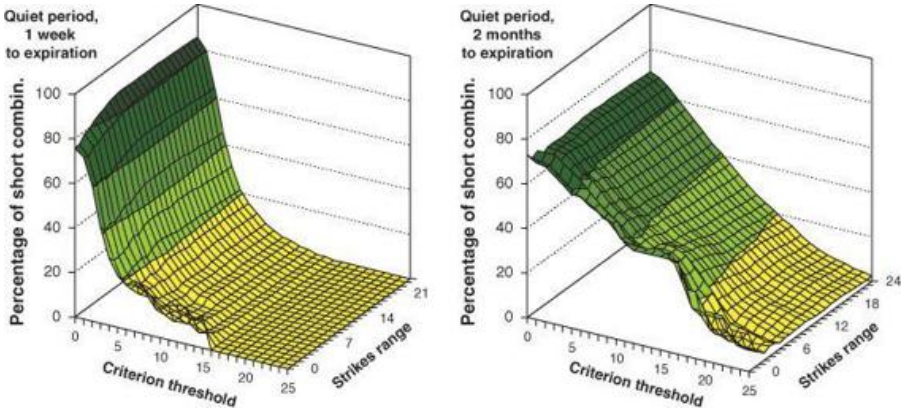

Figure 1.2.13. Dependence of the percentage of short combinations on the criterion threshold and strikes range. Two time periods from the time of portfolio creation until options expiration are shown.

In calm market conditions the percentage of short combinations depends only on the criterion threshold value. The breadth of strikes range has almost no influence on the characteristic under scrutiny (see Figure 1.2.13). To examine the relationship between the percentage of short combinations and the criterion threshold, and to compare different time periods left to options expiration, we reduced surfaces shown in Figure 1.2.13 to lines. It was accomplished by averaging data related to different strikes range values. Since the strikes range has no effect on the percentage of short combinations, this procedure will not lead to losses of information.

Averaged data presented in Figure 1.2.14 evidence that the percentage of short combinations in the portfolio reaches its maximum at low-criterion threshold values. In portfolios consisting of short-term options, this percentage is higher (more than 80%) than in portfolios constructed of options with a more distant expiration date (about 70%). The increase in criterion threshold leads to the decrease in the percentage of short combinations. If the portfolio consists of options with the nearest expiration date, the share of short combinations decreases at a much faster rate (exponentially). Furthermore, when the criterion threshold exceeds 15%, the percentage of short combinations decreases to zero. (This means that if the portfolio includes only those combinations for which the criterion value exceeds 15%, then short combinations fully disappear from the portfolio.) When the portfolio is composed of more distant options, the decrease in the percentage of short combinations is gradual (almost linear). At the criterion threshold value of 15%, more than a quarter of the portfolio still consists of short combinations.

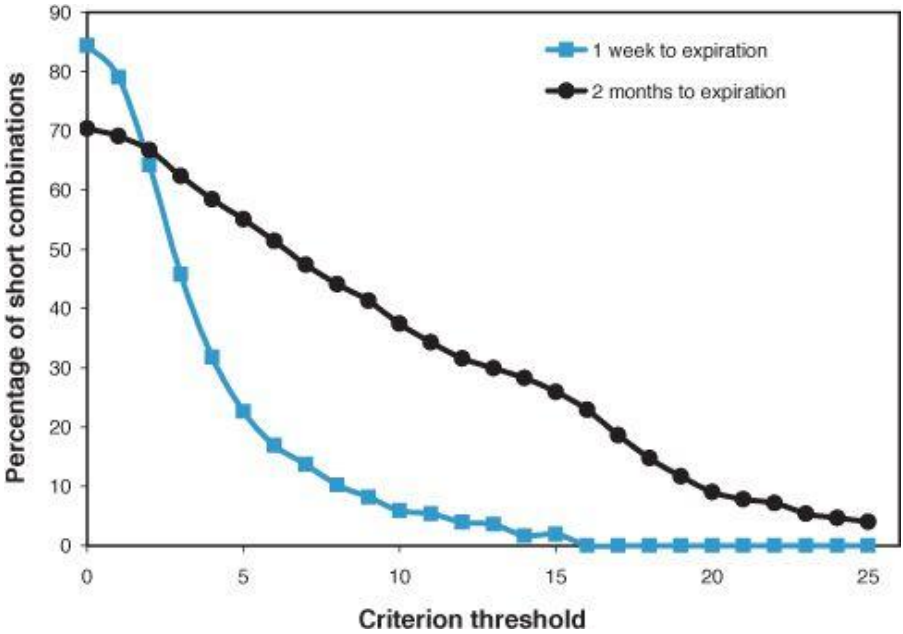

Figure 1.2.14. Relationship between the percentage of short combinations and the criterion threshold. Two time periods from the time of portfolio creation until options expiration are shown.

Proportions of Straddles and Strangles

The relative frequency of different types of option combinations in the portfolio does not depend on the criterion threshold value. On the other hand, the breadth of the strikes range influences the proportions of straddles and strangles significantly. A similar (but opposite) situation has been detected in the previous case, in which relative frequencies of long and short combinations did not depend on the strikes range, but were determined solely by the criterion threshold value. By analogy with that previous case, we have averaged data relating to different criterion threshold values. This enables us to better recognize the relationship between the percentage of straddles in the portfolio and the breadth of the strikes range, and to compare this relationship within different time intervals to options expiration and in different market conditions.

In all situations the share of straddles in the portfolio decreases as the strikes range increases (see Figure 1.2.15). This phenomenon can be explained by the fact that if only a narrow range of strike prices is allowed, then just a few strikes (or even only one strike) can get there. The fewer the strike prices that are available to create combinations, the fewer the strangles that can be constructed (and the higher the percentage of straddles in the portfolio). During calm markets, the decrease in the share of straddles is nonlinear—widening of the strikes

range up to 8% to 10% brings about a sharp and rapid decrease in the percentage of straddles. However, further widening of the range does not exert significant influence on this characteristic. When the strikes range is widened to 25%, the proportion of straddles falls to 8% (for portfolios consisting of two-month options) and 19% (for portfolios constructed of one-week options). During high-volatility periods, widening the strikes range leads to a smoother (almost linear) decrease in the share of straddles (see Figure 1.2.15). The time left to options expiration has almost no influence on the dynamics of this characteristic.

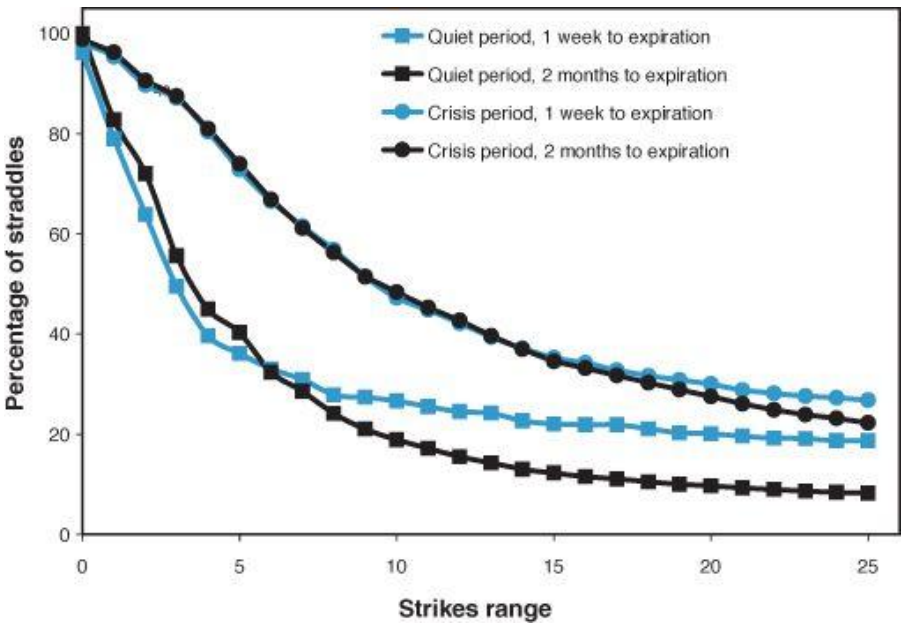

Figure 1.2.15. Relationship between the percentage of straddles and the strikes range. Two market situations (calm and crisis) and two time periods from the time of portfolio creation until options expiration (one-week and two-month) are shown.

Portfolio Asymmetry

This characteristic represents a special coefficient (described in detail in Chapter 3). It expresses the extent of asymmetry of the portfolio payoff function relative to the current value of a certain market index. Symmetry of a payoff function implies that the portfolio value will change by approximately the same amount regardless of whether the market is going up or down (if the extent of market growth and decline is similar). If the symmetry is violated, the payoff function is biased relative to the current index value, and the asymmetry coefficient reflects the degree of this bias. Since the concept underlying delta-neutral strategies is based on market

neutrality principles, the asymmetry coefficient is an important indicator that characterizes the balancing of the delta-neutral portfolio.

During calm markets, the asymmetry of portfolios created using the nearest options was rather low for almost all {criterion threshold × strikes range} combinations. However, at low-criterion threshold values and wide strikes range, the asymmetry coefficient turned out to be relatively high (the top-left chart in Figure 1.2.16). When long-term options are used to construct a portfolio, the chart surface changes from concave to convex (the bottom-left chart in Figure 1.2.16). This means that portfolio asymmetry gets higher at medium values of criterion threshold and strikes range. In such circumstances (calm market and long-term options) the symmetry persists in two areas: (1) at almost all criterion threshold intervals (given that the strikes range is rather narrow) and (2) at almost all strikes range values (given that the criterion threshold is high).

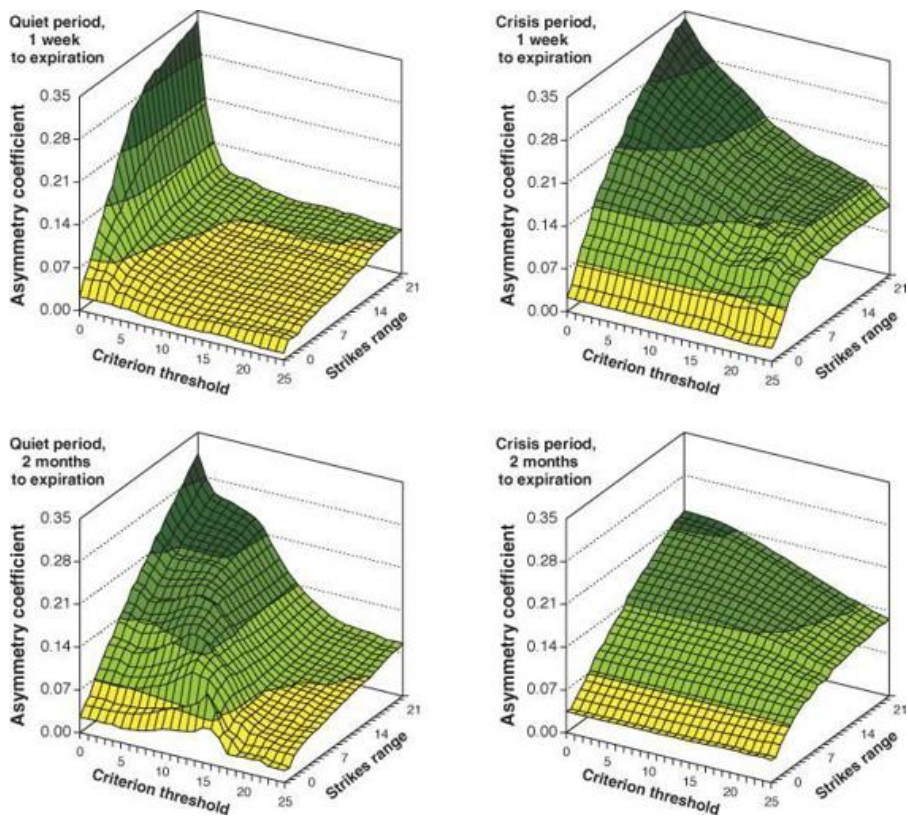

Figure 1.2.16. Dependence of the asymmetry coefficient on the criterion threshold and strikes range. Two market conditions (calm and crisis) and two time periods from the time of portfolio creation until options expiration (one-week and two-month) are

shown.

In periods of high volatility, the most asymmetrical portfolios are obtained when applying minimal criterion threshold values and maximal ranges of strike prices (the right-hand charts in Figure 1.2.16). However, while in calm times the asymmetry of portfolios consisting of the nearest options was generally lower than that of the portfolios constructed using long-term options, in volatile markets the picture is different. For all combinations of {criterion threshold × strikes range}, portfolios created using one-week options are less symmetric than portfolios constructed of two-month options (compare the top-right and bottom-right charts in Figure 1.2.16).

Loss Probability

This is an important characteristic expressing the risk of the option portfolio (its calculation methodology is described in Chapter 3). The top-left chart in Figure 1.2.17 relates to portfolios created in a calm market using the nearest options. The surface of this chart forms quite a broad plateau. This implies that loss probability remains monotonously high in wide intervals of criterion threshold and strikes range values. For portfolios consisting of combinations with criterion values exceeding 15% and strikes range wider than 18%, loss probability is about 60% to 65%. As the criterion threshold decreases and the strikes range narrows, loss probability diminishes. However, even under the most favorable combination of parameter values, loss probability does not fall below 40%.

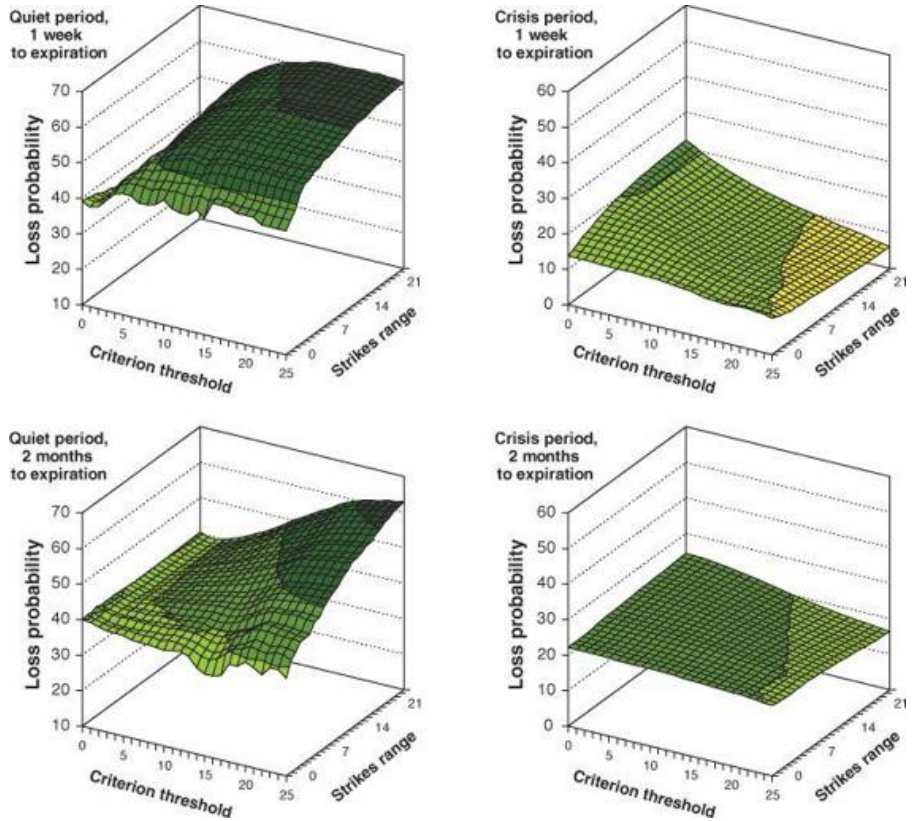

Figure 1.2.17. Dependence of the loss probability on the criterion threshold and strikes range. Two market conditions (calm and crisis) and two time periods from the time of portfolio creation until options expiration (one-week and two-month) are shown.

Comparison of the surface that has just been described with the chart related to portfolios composed of options with a distant expiration date (the bottom-left chart in Figure 1.2.17) reveals the following. Despite the general similarity of the two surfaces, there is one important difference between them. When long-term options are used to construct a portfolio, the wide plateau corresponding to the area of high loss probability turns into a small peak. This means that in such circumstances there are many fewer {criterion threshold × strikes range} combinations for which loss probability reaches its highest level. The descent from the area of the most probable loss is steeper than it was in the case of portfolios consisting of the nearest options. As a result, quite a broad area with relatively low loss probabilities is formed. This area covers criterion threshold values from 0% to 8% combined with strikes range values from 0% to 5%.

During crisis periods (the right charts in Figure 1.2.17), the loss

probability is significantly lower as compared to the calm market conditions. At first sight this seems illogical and suggests that there should be an error in probability estimates. However, this phenomenon can easily be explained by the fact that in an extreme market a portfolio consists primarily of short positions (this was discussed earlier in this section). Loss probability of short option combinations is much lower than that of long ones. This is evidenced by the data presented in Figure 1.2.18 (although this figure shows only one example, similar patterns were observed in all other cases that had been examined in the course of this study). It also follows from Figure 1.2.18 that loss probability of short combinations depends strongly on the criterion threshold value (the higher the threshold, the lower the loss probability), but does not depend on the strikes range. For long combinations the opposite relationship is observed: Loss probability is independent of the criterion threshold, but depends on the strikes range (the wider the range, the higher the loss probability).

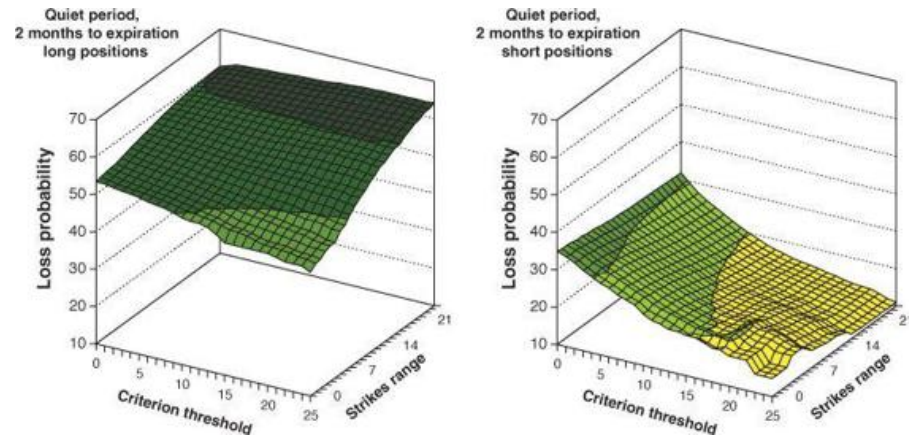

Figure 1.2.18. Dependence of the loss probability on the criterion threshold and strikes range as observed during calm market, when portfolios were constructed of options with two months left to expiration. Long and short positions are presented separately.

In a volatile market the loss probability of portfolios constructed of the short-term options reaches its maximum at low values of criterion threshold and at a wide range of strike prices (the top-right chart in Figure 1.2.17). This pattern differs from the picture observed in calm market conditions (the highest loss probability there was in line with high criterion threshold values; see the top-left chart in Figure 1.2.17). Loss probability of portfolios created during a volatile period using long-term options depends neither on criterion threshold nor on strikes range. The horizontal positioning and the flat shape of the surface of the bottom-right chart in Figure 1.2.17 confirm that.

This widely known risk indicator does not need special presentation, although its calculation for option portfolios has many specific peculiarities. Figure 1.2.19 shows that regardless of market conditions and irrespective of time left until options expiration, portfolio VaR is higher at lower values of criterion threshold. This holds true for almost all ranges of strike prices, although there are some exceptions. In particular, when portfolios are constructed of distant options (regardless of market volatility) and combinations are created using strikes situated in the narrow range around the current prices of underlying assets, the criterion threshold does not exert any influence on the VaR value.

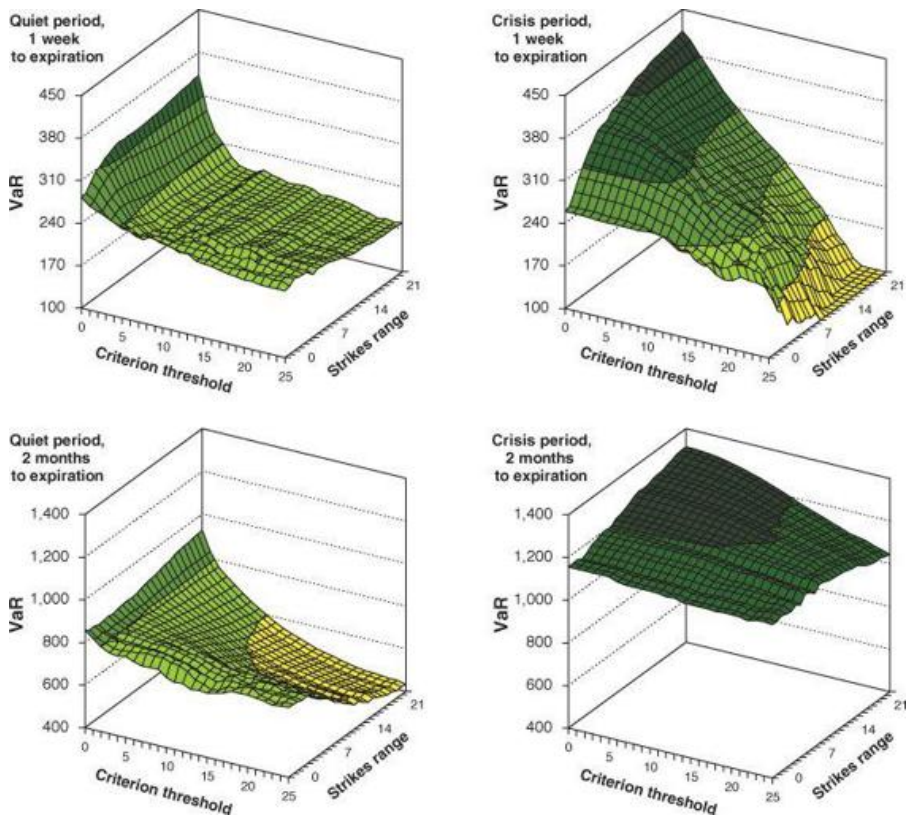

Figure 1.2.19. Dependence of VaR on the criterion threshold and strikes range. Two market conditions (calm and crisis) and two time periods from the time of portfolio creation until options expiration (one-week and two-month) are shown.

A comparison of portfolios constructed of the nearest options during the calm period (the top-left chart in Figure 1.2.19) and during an extreme market (the top-right chart in Figure 1.2.19) reveals that in the latter case VaR is slightly higher given the {low criterion threshold

× wide strike price range} combination. However, when the criterion threshold is high, VaR is lower during crisis than in calm market conditions. When portfolios consist of more distant options, the effect of market volatility becomes more distinct. Under all combinations of {criterion threshold × strikes range}, the VaR of portfolios created during the volatile period (the bottom-right chart in Figure 1.2.19) is higher than the VaR of portfolios created in the period of calm market (the bottom-left chart in Figure 1.2.19).

Surprisingly, the effect of the time period left to options expiration is much more pronounced than the effect of market volatility. When long-term options are used instead of short-term contracts, the VaR of all portfolios multiplies many times. This holds true for both calm and extreme market conditions.

1.3. Partially Directional Strategies

The main difference between partially directional and market-neutral strategies is that the forecast (predicting the direction and magnitude of future price movements) is among the basic elements of the former strategy. At the same time, the index delta (assessing the extent of market neutrality) also represents a constituent element of partially directional strategies. Although reducing delta to zero is not obligatory, observance of the delta-neutrality condition (or minimizing delta, if delta-neutrality is unattainable) while creating the option portfolio is highly desired.

1.3.1. Specific Features of Partially Directional Strategies

At first glance, simultaneous application of the two basic principles of the partially directional strategies—reliance on forecasts and adherence to market-neutrality—contradicts their inner logic. On the one hand, partially directional strategy is expected to benefit from changes in the underlying asset prices (if the forecast turns out to be true, the value of the option portfolio should increase as a result of price changes). On the other hand, efforts toward attaining the delta-neutrality imply that the strategy developer tends to create a portfolio that is insensitive to underlying asset price changes. This seeming contradiction disappears if forecasts are applied to separate combinations, while the index delta is used at the portfolio level. Applying forecasts to underlying assets of separate combinations leads to creating market-non-neutral (skewed) combinations. Nevertheless, combining these individually skewed combinations may lead to creation of a market-neutral portfolio. (If misbalances of separate combinations have opposite directions, unifying them should bring about construction of a market-neutral portfolio.)

Besides, the contradiction between using forecasts and striving for market-neutrality can be resolved naturally if we approach the problem of price movements differentially, depending on the level of price changes. Let us recall that reducing delta to zero minimizes the sensitivity of the portfolio to *small* changes in underlying asset prices. On the contrary, forecasts of price movements are oriented toward predicting *medium and large* price changes. Thus, minimization of delta ensures a portfolio insensitivity to small, chaotic, and mostly unpredictable price fluctuations. At the same time, using forecasts allows profiting from medium- and long-run price trends (if their directions have been forecasted correctly). This property differentiates partially directional strategies from the market-neutral ones, for which substantial price movements usually reduce a portfolio value regardless of the direction of the underlying asset price change.

There are various approaches to forecasting the future price of an underlying asset. In general, they can be divided into two main categories: forecasts based on technical analysis and forecasts based on fundamental analysis. Forecasts belonging to the first category use various technical indicators, statistical analysis, and probability theory. These methods rely, for the most part, on processing historical data relating to price, volume, volatility, and so forth. The second category includes forecasts built on the basis of macro- and microeconomic indicators. When stocks are used as underlying assets (most of the examples in this book use this class of underlyings), fundamental analysis relies on estimating market conditions, growth potential, and competition in the sector under scrutiny. It requires studying financial statements, considering different expert forecasts and many other information sources.

1.3.2. Embedding the Forecast into the Strategy Structure

Whichever methods of fundamental or technical analysis are used, the resulting forecast has to be quantifiable (that is, it must be expressed numerically). In particular, the forecast may be presented in one of the following forms:

- The most probable direction of future price movement
- The range of probable future prices
- Several price ranges with probabilities assigned to each of them
- Probabilities of each possible price outcome

In the last case the forecast represents a probability distribution, which is created by assigning the probabilities to the entire range of the discrete price series. (For continuous price such a forecast is

presented in the form of a probability density function.) This type of forecast is preferable since it contains the greatest volume of information relating to expected price movements of the underlying asset. In our further discussion we will use forecasts expressed in the form of either discrete distribution or probability density function.

There are several possible ways to embed the forecast into the strategy structure. Let us recall that signals for positions opening are generated on the basis of specially designed valuation criteria, most of which are integrals of the option combination payoff function over specific distribution. Hence, forecast incorporation into the procedure of the generation of trading signals can be done in one of two ways:

1. Adjustment of the discrete distribution (or probability density function) according to the forecast

2. Changing of the structure of the combination in a way that aligns the form of its payoff function with the forecast as much as possible

The combined application of these two methods is also feasible.

Let us consider the first method (embedding the forecast into the strategy by adjusting the distribution according to the forecast). For example, the strategy developer may rely on the forecast predicting that the underlying asset price will rise by 10%. Suppose that his trading system generates position opening signals by calculating the criterion based on lognormal distribution. In this distribution an expected price of the underlying asset is one of the two parameters. Earlier we mentioned that for a market-neutral strategy the value of this parameter is usually set to be equal to the current price as of the moment of criterion calculation. In the case of a partially directional strategy, the expected price may be estimated as the current price plus expected increment (10% in this case). This will shift the whole probability density function to the right (since the forecast predicts a price increase).

Embedding the forecast into the strategy structure by adjusting the distribution is shown in Figure 1.3.1. Black lines depict the probability density functions of lognormal distribution and the payoff function of the “butterfly” combination (which presumably has the highest criterion value as compared to other available option combinations). Embedding the forecast shifts the probability density function (the new function is shown by the light-colored line in Figure 1.3.1). Consequently, the criterion value is no longer maximal for the combination depicted by the broken black line. Now another combination (the light-colored dotted line) with the payoff function that fits the new distribution better has the highest criterion value.

Thus, *embedding the forecast by means of distribution adjustment leads to the selection of other combinations* (not those that would have been selected without the forecast).

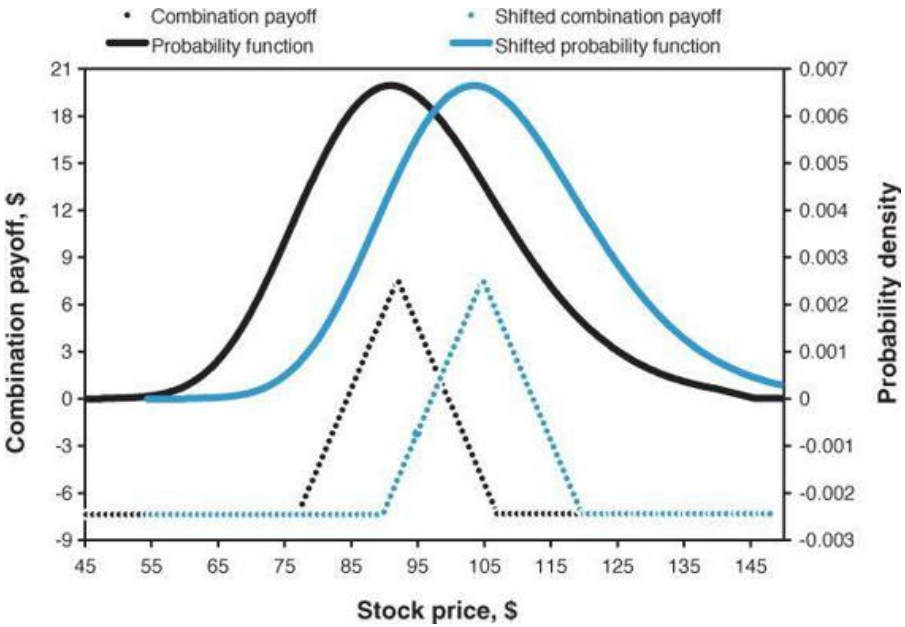

Figure 1.3.1. Probability density function of lognormal distribution and the payoff function of the combination with the highest criterion value (black lines). Embedding the forecast shifts the probability density function, which leads to the selection of another combination (light-colored lines).

The second method of embedding the forecast into the strategy consists in modifying the combination structure. Consider a forecast that indicates a high probability for the growth of the underlying asset price. For an opening short strangle (or straddle) position such a forecast may be embedded by creating the asymmetrical combination, wherein the number of call options sold is lower than the number of short puts. If the trading signal requires entering the long position, the number of call options bought should be higher than the number of long puts.

A great number of methods can be invented to embed the forecast into the strategy structure either by adjusting the distribution or by changing the combination structure (or by combining these two approaches). The examples used in this section to illustrate our discussion of partially directional strategies are based on the following principles. We construct the forecasts of future price movements on the basis of analysis of historical price series. For each underlying asset the probability of a certain price change is estimated by assessing the relative frequency of such outcomes observed during a predefined

time period (called “historic horizon”). After that, frequencies of past price movements obtained in this way are gathered into a discrete probability distribution (called “empirical distribution”), or they can be used to create a density function. Such an approach assumes that probabilities of different price movements can be predicted by the frequency of their past realizations. The properties of the empirical distribution and the method of its construction were described by Izraylevich and Tsudikman (2009, 2010).

In contrast to lognormal distribution, the probability density function of empirical distribution has an irregular shape, with numerous local peaks and bottoms, and in most cases is asymmetrical (see Figures 1.3.2 and 1.3.3). These irregularities reflect past price trends. For example, the skewing of distribution to the right or presence of local peaks on the right side of the distribution denote the prevalence of uptrends in the past. Thus, *presenting the forecast in the form of the empirical distribution can be regarded as an adjustment of standard lognormal distribution.*

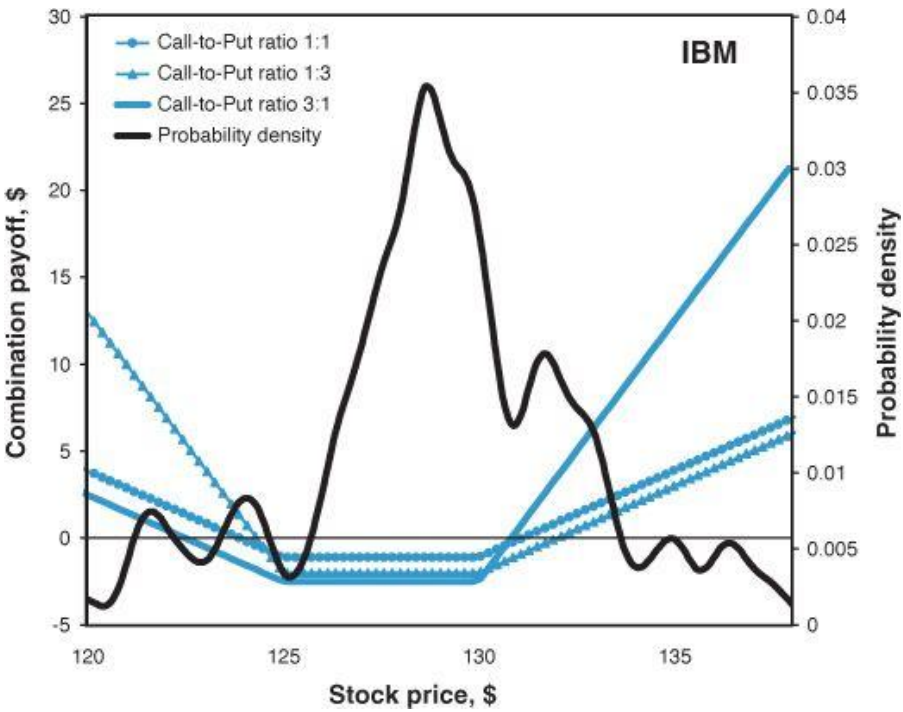

Figure 1.3.2. Probability density function of empirical distribution of IBM stock and payoff functions of three variants of the combination “long strangle.” The variants differ by the call-to-put ratio. The superior variant (having the highest criterion value) is shown with the bold light-colored line.

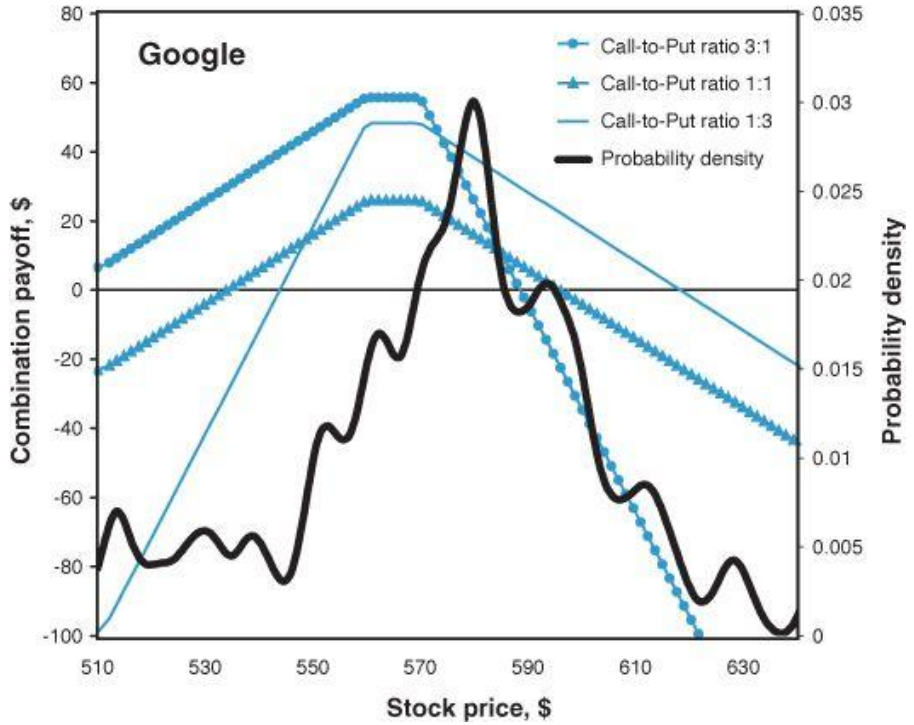

Figure 1.3.3. Probability density function of empirical distribution of Google stock and payoff functions of three variants of the combination "short strangle." The variants differ by the call-to-put ratio. The superior variant (having the highest criterion value) is shown with the light-colored line.

In addition to the distribution adjustment, we will also modify the combination structure (that is, a combined approach to embedding the forecast into the strategy will be implemented). Although a combination structure can be modified in several ways, we will use the simplest approach, which can be easily coupled with the empirical distribution—creating asymmetrical combinations. To apply this approach, several variants have to be created for each option combination. These variants will differ from each other by the relative ratio of call and put options (in the examples relating to a market-neutral strategy the equal ratio was used). The criterion value is calculated for each variant of this combination and the variant with the maximal criterion value is selected.

Let us consider several examples, in which the number of variants is limited to three call-to-put ratios: 3:1, 1:1, and 1:3. Suppose that we need to select the best variant of the long strangle for IBM stock. Figure 1.3.2 shows the probability density function of empirical distribution (constructed on the basis of a 120-day historical period)

and payoff functions for three variants of this combination (constructed using strike prices of put 125 and call 130). The distribution was created on April 1, 2010. The expiration date of the options is April 16, 2010. In this example the highest value of "expected profit on the basis of empirical distribution" criterion was obtained for the combination consisting of three calls and one put. The superiority of this particular variant can be explained by the shape of empirical distribution, which is skewed to the right (besides, there is a local peak on the right side of this distribution). Such a shape of the probability density function forecasts the increase of the underlying asset price. Hence, the long combination, in which the number of call options exceeds the number of puts, has higher profit potential (if the forecast turns out to be correct).

In the next example we used Google stock as the underlying asset. Suppose that this time the best variant has to be selected for another type of option combination, the short strangle. The probability density function of empirical distribution and payoff functions of three strangle variants are shown in Figure 1.3.3. As in the preceding example the distribution was constructed on April 1, 2010. All three combination variants were created of options expiring on April 16, 2010 (using strike prices of put 560 and call 570). Although the appearance of empirical distribution differs from the shape of distribution observed in the preceding example (compare Figures 1.3.2 and 1.3.3), it is also skewed to the right. This means that the forecast embedded into the partially directional strategy indicates a high probability of the underlying asset price appreciation. In such circumstances it would be reasonable to sell more put options. If the forecast comes true, the stock price will rise and the loss incurred by the short call will be lower (because this option was sold in less quantity). At the same time, short puts will most likely expire out-of-the-money and thereby will produce higher profit (since they were sold in greater quantity). In full compliance with this logic, the highest criterion value is obtained for the combination variant with a 1:3 call-to-put ratio.

1.3.3. The Call-to-Put Ratio at the Portfolio Level

Up to this point we considered the relative frequencies of call and put options at the level of separate combinations. When the call-to-put ratio deviates from 1:1, the payoff function of the combination becomes asymmetrical (see Figures 1.3.2 and 1.3.3). Unifying several combinations may lead to noticeable transformation in the shape of the portfolio payoff function. The resulting changes in the payoff form can be very different.

At the portfolio level the payoff function expresses the relationship between portfolio profits/losses and a certain market index. To create such a function, we need to establish the relationships between the index value and profits/losses of all individual combinations composing the option portfolio. This can be done using the concept of beta (this issue is discussed in detail in Chapter 3). The payoff functions of separate option combinations constructed by applying this technique are additive. Their mere summation produces the payoff function of the resulting complex portfolio.

Unifying homogenous option combinations (that is, combinations of the same type with the same call-to-put ratio) gives rise to the payoff function that resembles the functions of input combinations. The light-colored lines in the top-left chart of Figure 1.3.4 show the payoff functions of two short strangles with the call-to-put ratio of 1:3. Such a ratio causes the left parts of both functions to become steeper as compared to their right parts (this implies that the decrease in the index value results in higher losses than its increase). The payoff function of the portfolio (black line) consisting of these two combinations has a similar shape.

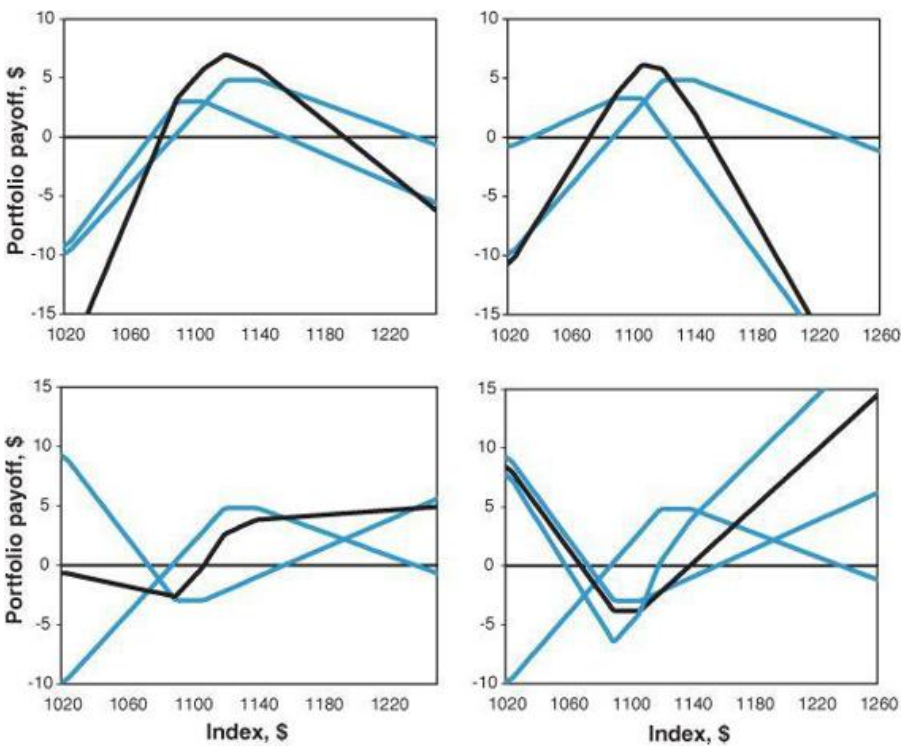

Figure 1.3.4. Effect of the combinations asymmetry (resulting from unequal call-to-put ratios) and of the ratio of long and short combinations on the payoff function of the resulting portfolio.

Payoff functions of separate combinations are shown by light-colored lines; payoff functions of resulting portfolios are shown by black lines.

If the call-to-put ratios of input combinations are different, the characteristics of the resulting portfolio, including its payoff function, will depend on the weighted average of the initial ratios. For example, unifying two short strangles having the call-to-put ratios of 3:1 and 2:3 leads to the creation of a portfolio with a 5:4 ratio. The light-colored lines in the top-right chart of Figure 1.3.4 show two short strangles with ratios of 1:3 and 3:1. Both combinations are asymmetrical but they are skewed to different sides. Unifying such combinations results in an almost symmetrical payoff function (black line in the chart). Within the partially directional strategy this approach enables creating the option portfolio, in which every element is an asymmetrical combination, but nevertheless the whole portfolio remains market-neutral (or approaches the market-neutrality closely).

Earlier we stated that the relative ratio of long and short combinations significantly influences its main properties. When asymmetric combinations are allowed, this statement takes on a special meaning. Different call-to-put ratios in conjunction with different proportions of long and short combinations allow for creating a great variety of shapes of portfolio payoff functions (sometimes they assume very intricate forms). For example, unifying long and short strangles with the call-to-put ratio of 1:3 generates the payoff function resembling that of the “bull spread” combination (the bottom-left chart of Figure 1.3.4). At the same time, adding more long or short combinations (thereby amending the resulting ratio of long and short combinations) may bring about the fundamental change in the shape of the portfolio payoff function. In the example presented in the bottom-left chart of Figure 1.3.4, introduction of one additional long combination with call-to-put ratio of 2:3 transforms the portfolio payoff function and makes it similar to the typical long straddle (the bottom-right chart of Figure 1.3.4).

1.3.4. Basic Partially Directional Strategy

Signals for positions opening and closing. Signals to open trading positions are generated by evaluating option combinations according to the values of the “expected profit on the basis of empirical distribution” criterion. As mentioned earlier, using empirical distribution is regarded in the context of partially directional strategy as one of the two ways to embed the forecast into the strategy structure. (The second way, which is also used in the basic variant of

this strategy, consists of creating asymmetric combinations.) The opening signal is generated when the criterion value exceeds a certain threshold value. The range of permissible values for the threshold parameter spans from zero to infinity. The exact threshold value is determined via optimization. Under the basic variant of partially directional strategy, open positions are hold until options expiration. After the expiration, all positions in underlying assets resulting from the exercise of options are closed the next trading day.

Criterion parameters. The algorithm for calculation of the "expected profit on the basis of empirical distribution" criterion was described by Izraylevich and Tsudikman (2010). In contrast to the lognormal distribution, the empirical one is based on a single parameter: the history horizon (the length of the price series that serves to construct the distribution). As in the case of the basic market-neutral strategy, we set the value of the history horizon parameter at 120 days. The forecast horizon (number of days from the current time until the future date for which the distribution is constructed) is determined automatically by fixing the date for which the criterion is calculated.

Selection of option combinations. The initial set of underlying assets available to create option combinations includes all stocks constituting the S&P 500 index. The acceptable types of option combinations include long and short strangles and straddles. Five variants of call-to-put ratios (one symmetric and four asymmetric) are created for each combination: 3:1, 3:2, 1:1, 2:3, 1:3. For each variant the criterion value is calculated, and then the variant with the highest value is selected. If the criterion value obtained for the best variant exceeds the threshold value, the opening signal for this combination is generated. Under this approach the resulting ratio of calls and puts is not set a priori, but depends on ratios of separate combinations included in the portfolio. Hence, the extent of the portfolio divergence from market-neutrality is not set a priori by the developer, and the index delta is not used at the stage of portfolio formation (only for risk management).

Capital management and allocation within the investment portfolio. In a basic variant of the partially directional strategy, all available funds are invested in the portfolio. All funds released after positions closing are reinvested. Capital distribution between portfolio elements is based on the principle of equivalence of stock positions (the position volume is set in such a manner that should the options be exercised, the amount of funds invested in all underlying assets will be approximately equal; see Chapter 4 for details). Applying this principle to a partially directional strategy is complicated by the

inequality of the call and put quantities within a given combination. For asymmetrical combinations the amount of funds required to fulfill the obligation (or to exercise the right) depends on which side of the combination, either call or put, is exercised (that is, which of the options is in-the-money). If unused capital at some moment in time equals C, and m opening signals have been generated at that time, the position volume for each combination is

$$N_c = \frac{4C \times r}{m \times (S_c + S_p) \times (1+r)}, \quad N_p = \frac{4C}{m \times (S_c + S_p) \times (1+r)}$$

where r is the call-to-put ratio, N_p and N_c are the quantities of call and put options included in the combination, and S_c and S_p are the strike prices of call and put options, respectively.

Risk management. In contrast to market-neutral strategies, adherence to the principle of delta-neutrality is not a compulsory condition in the risk management of partially directional strategy. Nevertheless, the index delta, in this case too, remains the main instrument of risk assessment and control. This indicator evaluates potential losses that may be incurred under adverse market conditions. The index delta also expresses (indirectly) the extent of portfolio disequilibrium and the degree of its payoff function asymmetry. Other indicators (VaR, loss probability, and asymmetry coefficient) are also useful in estimating and managing the risks of a partially directional strategy.

1.3.5. Factors Influencing the Call-to-Put Ratio in an Options Portfolio

The call-to-put ratio strongly influences the shape of the portfolio payoff function. Various factors simultaneously influence the call-to-put ratio in the portfolio. The effects of some of them are illustrated here for the basic partially directional strategy. The statistical research presented next is based on the database containing prices of options and their underlying assets from March 2000 until April 2010. During this period, we simulated option portfolios according to the principles stated for the basic strategy. The “criterion threshold” parameter was set at zero (that is, the opening signals are generated for all combinations with a positive criterion value), and the “strikes range” parameter was set at 50%.

The relative frequency of call and put options in the portfolio may be expressed in three ways. This can be illustrated using the data presented in the bottom-right chart of Figure 1.3.4. This portfolio is composed of one short combination with a 1:3 call-to-put ratio and two long combinations with 1:3 and 2:3 ratios. *The first method*

consists of a summation of all call and put options. In our example this gives the ratio of 4:9 (or 0.44). This method does not account for the fact that the influence of call options included in the short combination on the payoff function is neutralized to some extent by the influence of calls belonging to the long combinations. (The same is true regarding put options.) *The second method* consists of presenting the call-to-put ratios of short combinations as negative numbers. Application of this principle to our example produces the ratio of 2:3 (or 0.67) because the short combination with the 1:3 ratio and the long combination with the same ratio offset each other completely. Under certain conditions the call-to-put ratio can be negative. If the portfolio consists of many long and short combinations, their mutual offsetting can distort the information about portfolio call-to-put ratio. This represents an essential drawback of the second approach. For example, if the portfolio consists of ten short strangles with a 1:1 ratio, ten long strangles with the same ratio, and one long strangle with a 1:2 ratio, the final call-to-put ratio is 1:2 (0.5). Does this value reflect the real situation? It's quite doubtful.

In our opinion, it is preferable to use the *third method* of expressing the relative frequency of calls and puts in the option portfolio. This method consists of calculating the call-to-put ratio separately for long and short positions. The percentage of long and short combinations in the portfolio should also be taken into account. This approach enables expressing the effects of individual combinations on the portfolio structure and on its payoff functions more accurately. The situation is complicated, however, by the fact that these two characteristics (the call-to-put ratio and the percentage of short combinations) are interrelated.

We begin the exploration of factors influencing the call-to-put ratio with the percentage of short combinations in the portfolio. Note that this characteristic is not really a "factor" as we generally understand this term. Speaking about factors influencing the characteristic under scrutiny, we usually imply some variable that is external with respect to the system under investigation and, in most cases, independent of it. In the basic variant of partially directional strategy, the percentage of long and short positions is not an independent variable. It is determined automatically in the process of portfolio creation and depends on the algorithms used to generate opening signals and on different parameters of the trading strategy. Besides, the percentage of short combinations can also be affected by market conditions prevailing at the time of portfolio creation. Thus, it would be more correct to discuss the *interrelationship* between the percentage of short combinations and the call-to-put ratio rather than the *influence* of the

former characteristic on the latter one.

It follows from Figure 1.3.5 that the higher the percentage of short combinations in the portfolio, the higher the call-to-put ratio in these short positions. In cases when short combinations comprised less than 60% to 70%, they were significantly skewed toward the put side (puts outnumbered call options). However, the situation changed when short combinations comprised a higher percentage. In these cases calls became a majority within short positions. The relationship was the contrary in long positions: The higher the proportion of short combinations in the portfolio, the lower the call-to-put ratio in long combinations. When shares of long and short positions were approximately equal (50% of short combinations), long combinations consisted primarily of calls (1.5 times more calls than puts), while short combinations were mainly constructed of puts (2 times more puts than calls).

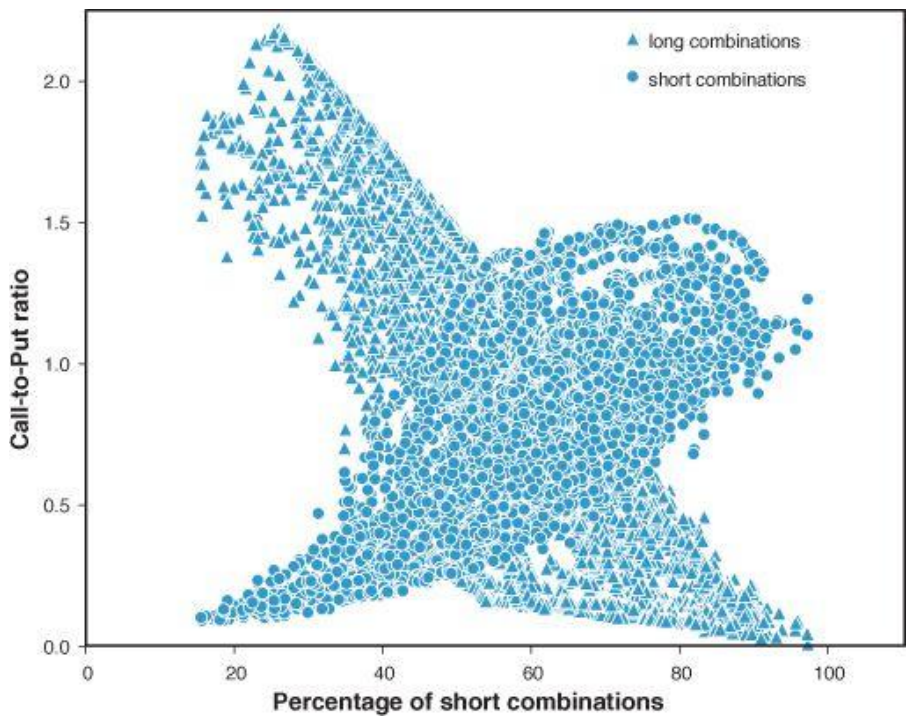

Figure 1.3.5. Relationship between the call-to-put ratio and the percentage of short combinations in the portfolio. The data (presented separately for long and short positions) relates to a ten-year historical period.

To comprehend the implications of the relationships presented in Figure 1.3.5 for the basic partly directional strategy, we need to consider another relationship. Reviewing the properties of the basic

delta-neutral strategy, we noted that in periods of high volatility all portfolios consisted of almost only short positions. For a partly directional strategy we also discovered the positive relationship between the percentage of short combinations and market volatility (see Figure 1.3.6). This relationship was quite weak for portfolios created close to options expiration ($R2 = 0.06$), but became stronger as the time left until expiration increased ($R2 = 0.18$ for a two- to three-month period, $R2 = 0.30$ for a four- to six-month period).

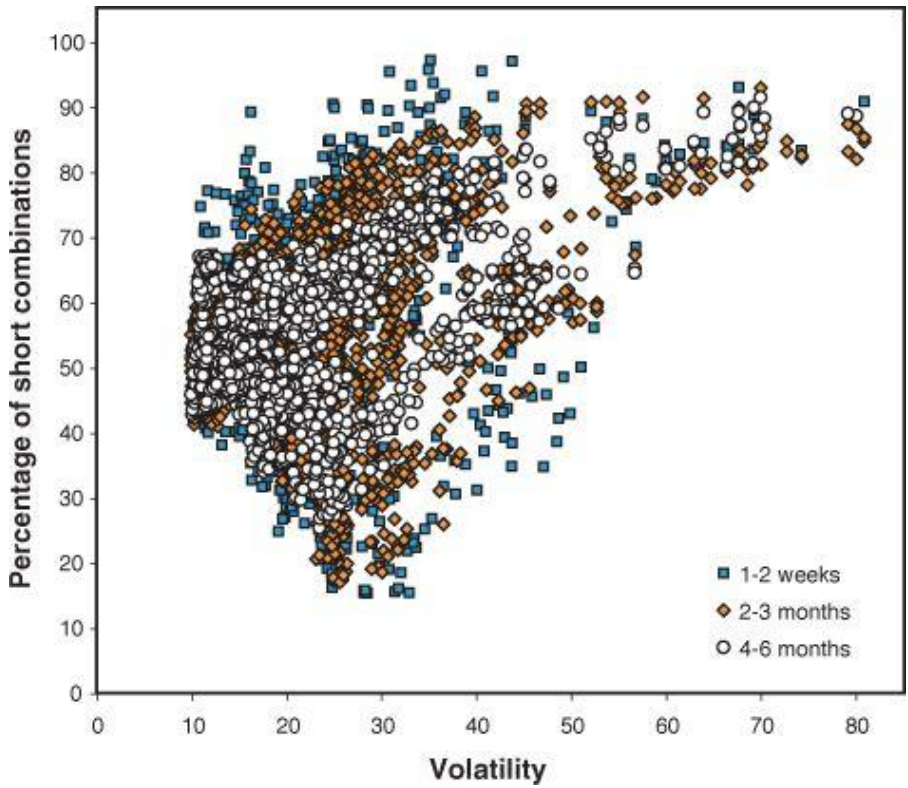

Figure 1.3.6. Relationships between the percentage of short combinations in the portfolio and the volatility of the underlying assets market. Three periods from the time of portfolio creation until options expiration are shown. The data relates to a ten-year historical period.

Tracing the relationships shown in Figures 1.3.5 and 1.3.6 brings about the following chain of reasoning. During crisis periods, when market volatility increases significantly, the criterion generates signals to open mainly short positions. (It happens due to the sharp rising of option premiums in extreme markets.) Basically, this seems quite risky since during a crisis there is a high probability of great price drops that generally cause significant losses to short option positions. However, in the case of partially directional strategy, this risk is

largely offset by predominance of call options in short combinations during periods of high volatility. Since in a crisis period the risk of a sharp price increase is lower than the decline risk, premiums obtained from the sale of extra call options may compensate, at least partially, potential losses occurring in the case of unfavorable developments taking the market down. In long positions (in crisis periods there are few of them, but still there are some), the number of puts is higher than the number of calls. This also diminishes the total risk of the portfolio because in the case of a market breakdown, excess put options yield additional profit which offsets partially the losses incurred by short positions.

These arguments suggest that *the structure of the partially directional strategy contains a certain built-in mechanism providing for the skewness of combinations payoff functions in response to changes in market conditions* (volatility jumps). To verify this supposition, we have to analyze (directly) the relationship between the call-to-put ratio and market volatility. (Until now we examined only the indirect effect of volatility. It was shown that volatility may influence the relative frequency of call and put options through its influence on the percentage of short combinations.) For long combinations we detected the inverse nonlinear relationship between the call-to-put ratio and market volatility (the left chart in Figure 1.3.7). It should be noted that at low volatility levels (relating to the calm market conditions) the ratio under consideration fluctuated in a quite wide range of values. However, during crisis periods (at high volatility levels) the quantity of puts in long combinations always outnumbered the call quantity (this ensures functioning of mechanisms allowing for the reduction of risks of short positions). As expected, a direct (also nonlinear) relationship between the call-to-put ratio and volatility was discovered for short combinations (the right chart in Figure 1.3.7). At low volatility levels the ratio varies in a wide range, but usually does not exceed one (this means that short combinations are skewed toward a put side). However, volatility growth leads to a sharp increase in the call-to-put ratio. At extreme volatility values this ratio exceeds one in most cases. Thus, current study provides rather strong reasons to state that *the call-to-put ratio, changeable depending on market conditions, represents an automatic (that is, launching automatically during a crisis) regulator of the option portfolio risk.*

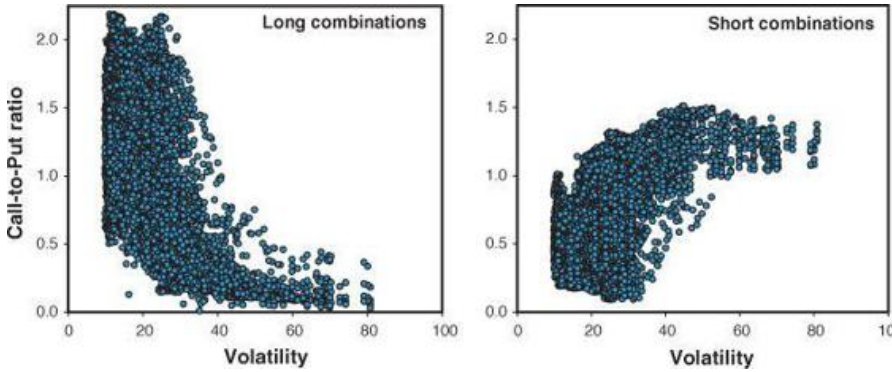

Figure 1.3.7. Relationship between the call-to-put ratio and the volatility of the underlying assets market. The data (presented separately for long and short positions) relates to a ten-year historical period.

In conclusion of this study, let us consider the influence exerted by another important factor, the number of days left until options expiration. The effect of this factor on various aspects of a delta-neutral strategy was repeatedly shown in the preceding section. Its influence is not less pronounced in the case of a partially directional strategy. Close to the expiration date, the call-to-put ratio is significantly higher in long than in short positions (see Figure 1.3.8). While long combinations during this period were on average symmetric (the call-to-put ratio approximately equals 1), in short combinations the number of put options was two times higher than the number of calls. Such a structure of short combinations, providing higher profit potential in a growing market, reflects the bullish trends that have prevailed in the stock market over the past ten years. (Recall that in the basic variant of partially directional strategy, the forecast is embedded into the strategy by the use of empirical distribution that reflects trends prevailing in the market in the past.) Even the last financial crisis has not changed the bullish skewness of an average short combination constructed of options with the nearest expiration date, though it undoubtedly has had an effect on variability of the call-to-put ratio.

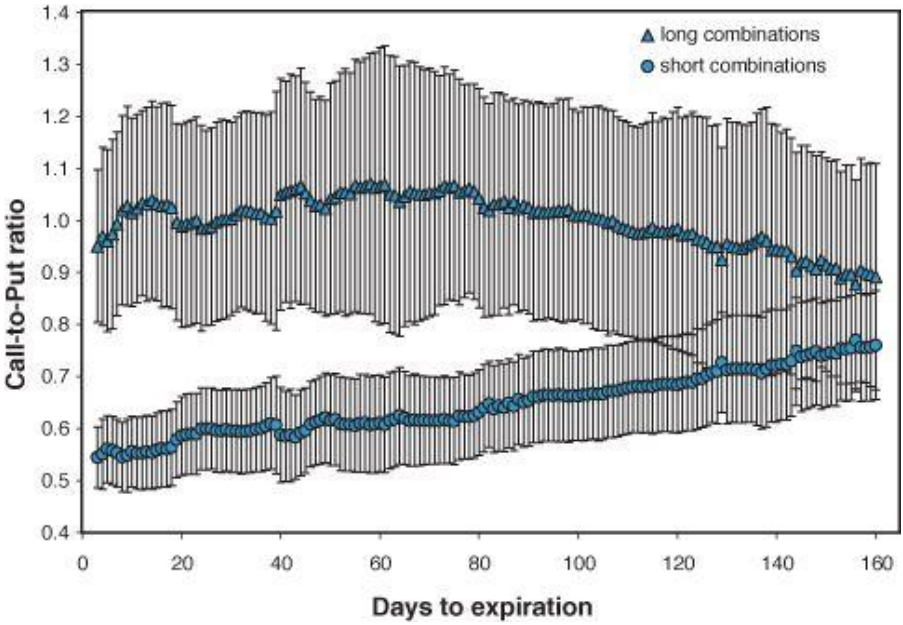

Figure 1.3.8. Relationship between the call-to-put ratio (mean ± standard deviation) and the time left until options expiration. The data (presented separately for long and short positions) relates to a ten-year historical period.

The more that distant options are used to create combinations, the less the difference is in the structure of long and short positions. Figure 1.3.8 distinctly shows two opposite trends: Whereas in long combinations the call-to-put ratio decreases, in short combinations this characteristic grows as the time to expiration increases. Besides, it is noticeable that the variability of the call-to-put ratio (shown in Figure 1.3.8 with vertical bars reflecting standard deviations) is significantly higher in long combinations than in short ones. This difference holds for the whole range of time periods left to option expiration. All these relationships are of a great importance for the optimization of parameters of a partially directional strategy.

1.3.6. The Concept of Delta-Neutrality as Applied to a Partially Directional Strategy

Although the observance of the delta-neutrality principle is not an obligatory condition (the key element of the strategy is the forecast of underlying assets prices), minimization of the index delta represents the second-most-important component of the partially directional strategy. As a result of forecast application, call-to-put ratios of individual combinations deviate from parity, which leads to an asymmetry of their payoff functions. By definition, such combinations cannot be market-neutral. Nevertheless, the portfolio consisting of

different asymmetric combinations may be delta-neutral (if the index deltas of separate combinations have different signs, their sum may be equal to or close to zero). Although the developer of a partially directional strategy should strive for effective embedding of the forecast into the strategy structure, at the same time he needs to minimize the portfolio index delta as much as possible. A reasonable compromise is required here; its effectiveness determines, to a large extent, the quality of the strategy.

Considering the relationship between the relative frequency of call and put options in the portfolio and its index delta reveals the following. In those cases when the call-to-put ratio is close to the parity, the index delta of short combinations diverges slightly from zero toward negative values and the index delta of long combinations deviates in the opposite direction, toward positive values (see Figure 1.3.9). Hence, when the call-to-put ratio is close to one (for both long and short positions) and the proportions of long and short positions are approximately equal, grouping all combinations should lead to the creation of a delta-neutral portfolio. However, the conditions under which the delta-neutrality of a partially directional portfolio may be attained are not limited to these circumstances.

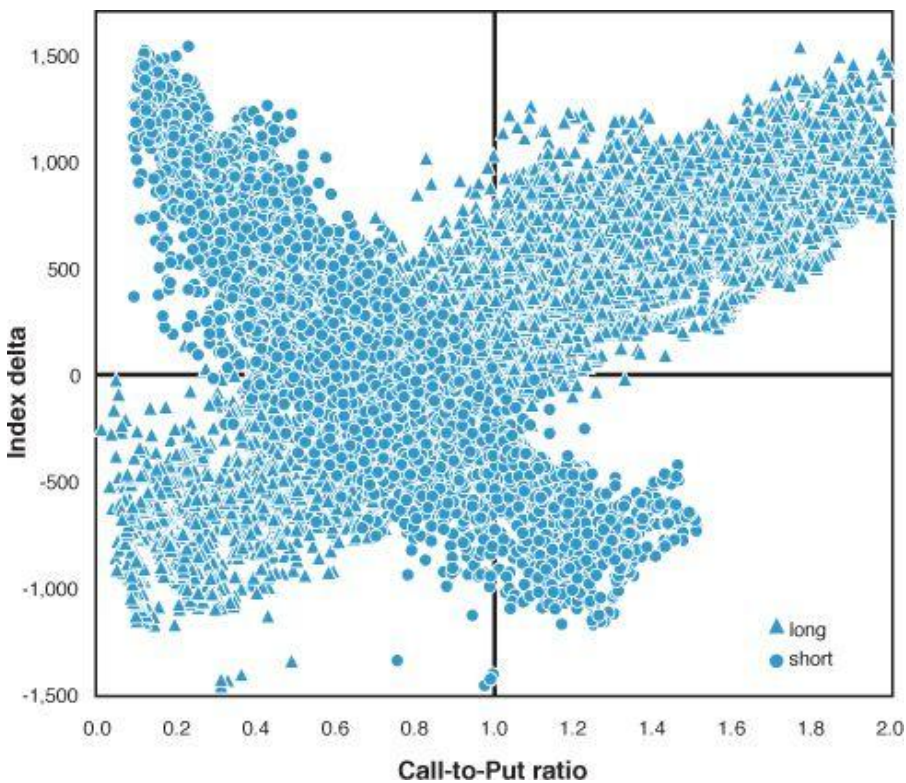

Figure 1.3.9. Relationship between the portfolio index delta

and the call-to-put ratio. The data (presented separately for long and short positions) relates to a ten-year historical period.

We detected a direct relationship between the call-to-put ratio in long combinations and the index delta of the portfolio. While the prevalence of put options (call-to-put < 1) leads to negative delta, the predominance of calls (call-to-put > 1) brings it to a positive area (see Figure 1.3.9). For short combinations this relationship is opposite: there is an inverse relationship between the call-to-put ratio and the index delta. Puts predominance leads to positive delta; calls prevalence, to negative delta. The opposite directions of these relationships suggest that combining several portfolio variants (with different call-to-put ratios) may lead to delta-neutrality of the global portfolio (despite significant deviations from zero in deltas of separate portfolios).

In the following text we present the extensive analysis of delta-neutrality boundaries similar to the one shown earlier for a delta-neutral strategy. Two samples, relating to periods of low and high volatility, were taken from the ten-year database. Within each sample two portfolio variants were created: portfolios consisting of close options (expiring in one to two weeks) and those constructed of more distant contracts (two to three months until expiration). Each delta-neutrality boundary is described by four quantitative characteristics: *criterion threshold index* (determines the position of the delta-neutrality boundary relative to the range of allowable values of the "criterion threshold" parameter); *strikes range index* (determines the position of the delta-neutrality boundary relative to the range of allowable values of the "strikes range" parameter); *length of delta-neutrality boundary* (expresses the number of points composing the delta-neutrality boundary); and *delta-neutrality attainability* (expresses the probability of creating delta-neutral portfolios).

Figure 1.3.10 shows criterion threshold and strikes range indices. Visual analysis of these data brings us to the following conclusions:

- *During calm market periods, delta-neutrality boundaries are situated in a quite wide range of criterion threshold values.* This conclusion is based on the positioning of filled circles and triangles along the horizontal axis of Figure 1.3.10.
- *During crisis periods, delta-neutrality boundaries are less variable by the criterion threshold value than during low-volatility periods.* This conclusion is based on the predominance of contour circles and triangles along the left part of the horizontal axis of Figure 1.3.10.
- *Delta-neutrality boundaries of portfolios consisting of options*

with the nearest expiration date are located in the area of lower criterion threshold values than boundaries of portfolios constructed of distant options. This conclusion (which is accurate for both calm and volatile markets) is based on the positioning of circles to the left of the triangles with respect to the horizontal axis of Figure 1.3.10.

• *The time left until options expiration influences the location of delta-neutrality boundaries with respect to the strikes range axis. Boundaries of portfolios consisting of the nearest options are situated in the area of lower strikes range values than boundaries of portfolios formed of distant options.* This conclusion (which holds true for both calm and volatile markets) is based on the lower position of the triangles as compared to the circles with respect to the vertical axis of Figure 1.3.10.

• *As compared to calm market conditions, during volatile periods delta-neutrality boundaries are located in the area of higher strikes range values.* This conclusion (which is accurate for portfolios constructed of both nearest and more distant options) is based on the fact that filled circles and triangles are located lower than the contour signs with respect to the vertical axis of Figure 1.3.10.

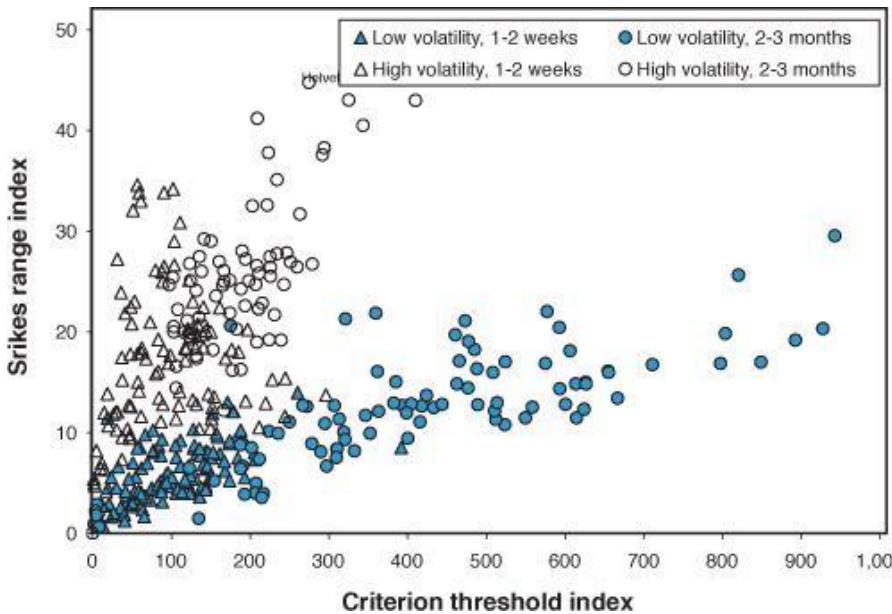

Figure 1.3.10. Presentation of delta-neutrality boundaries as single points with coordinates matching the criterion threshold and strikes range indexes. The data relates to a ten-year historical period.

The conclusions drawn earlier from a similar analysis of the data relating to the delta-neutral strategy (see Figure 1.2.7) differ in many respects from the findings of this study. This means that the effects of factors determining the positions of delta-neutrality boundaries may vary depending on the class of the option strategy under consideration. Hence, *simultaneous application of several heterogeneous strategies enables creating an automated trading system with a high likelihood of constructing and maintaining portfolios that are globally delta-neutral regardless of external circumstances* (that is, for almost all possible combinations of influencing factors).

Now we turn to the issue of delta-neutrality boundaries length. This characteristic is quite important since longer boundaries enable creating more variants of a delta-neutral portfolio (by manipulating values of the "criterion threshold" and "strikes range" parameters). The dependence of boundary length on criterion threshold and strikes range indexes is shown in Figure 1.3.11. The longest boundaries are situated in the area of high criterion threshold values and average values of strikes range. This means that *to obtain a longer delta-neutrality boundary, it is necessary to select combinations with high criteria values and to use an intermediate range of strike prices.* The lowering of the criterion threshold value leads to shortening of delta-neutrality boundaries. The same effect holds for narrowing or widening the strikes range.

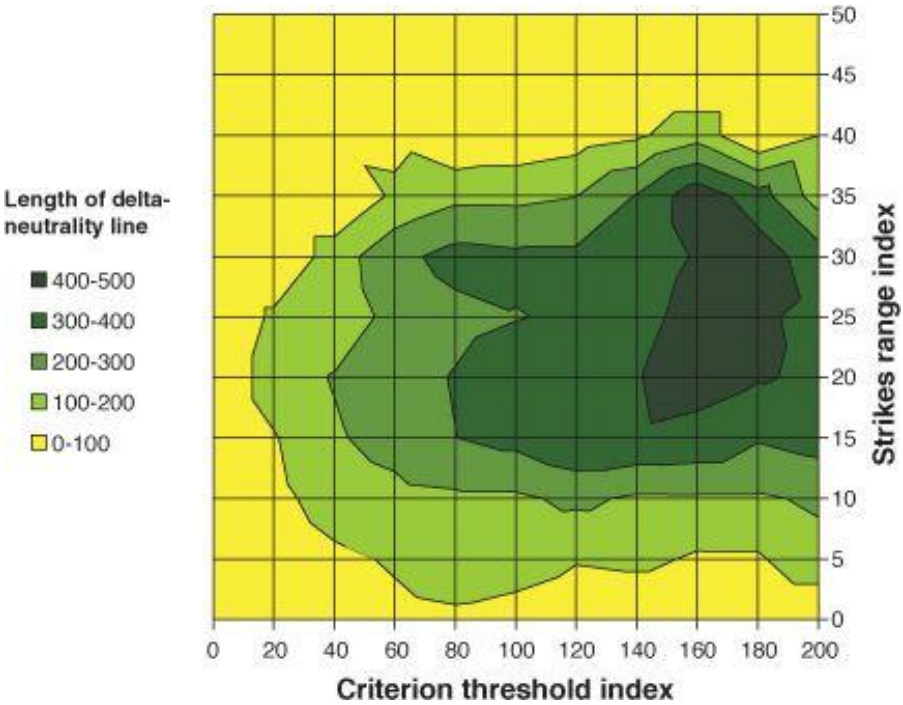

Figure 1.3.11. Relationships between the length of the delta-neutrality boundary and the indexes of criterion threshold and strikes range. The data relates to a ten-year historical period.

Parallel consideration of the data presented in Figures 1.3.11 and 1.2.8 allows comparison of two option strategies in terms of the potential effects of two parameters (criterion threshold and strikes range) on the length of delta-neutrality boundaries. Basically, the topographic maps shown in these two figures look quite similar. The only difference consists in the effect of the strikes range. For the delta-neutral strategy the longest boundary length is achieved under a narrow strikes range, whereas for the partially directional strategy the maximization of the boundary dimension requires using an average range of strike prices.

The dependence of the delta-neutrality boundary length on market volatility and time left to options expiration is presented in Figure 1.3.12. The longest boundaries are observed under the conditions of average volatility and short time to expiration. If portfolios are constructed using options with less than 10 days left until expiration, the borderline shortens. Using long-term options (more than 20 days to expiration) leads to a similar result. Regardless of the time remaining to options expiration, the delta-neutrality boundary shortens when the volatility deviates from its average values (either increases or decreases).

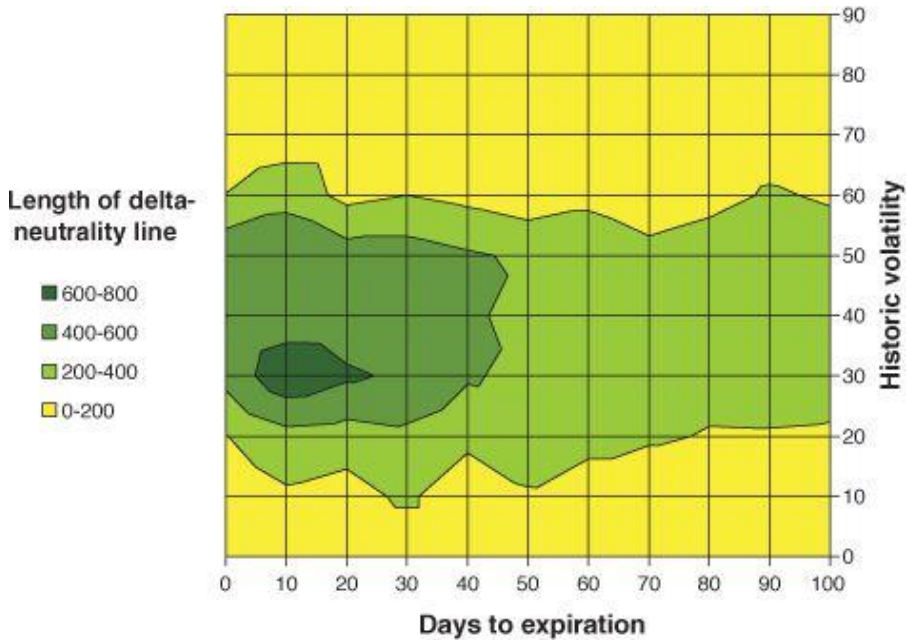

Figure 1.3.12. Relationships between the length of the delta-neutrality boundary, the number of days left until options

expiration, and market volatility. The data relates to a ten-year historical period.

A comparison of Figures 1.3.12 and 1.2.9 in terms of the influence of volatility and time to expiration on the length of delta-neutrality boundaries reveals a slight difference between the two option strategies. The main dissimilarity between the topographic maps presented in these two figures lies in the volatility effect. For the delta-neutral strategy the boundary length reaches its highest values under high-volatility conditions, whereas for the partially directional strategy the longest boundaries go with moderate volatility levels. In other respects the effects of volatility and time to expiration do not differ between the two option strategies. Thus, although the positions of delta-neutrality boundaries are quite specific for delta-neutral and partially directional option strategies (compare Figures 1.3.10 and 1.2.7), these strategies are similar with regard to the boundaries length (compare Figure 1.3.11 with Figure 1.2.8, and Figure 1.3.12 with Figure 1.2.9).

Let us examine the extent to which the delta-neutrality is attainable for the partially directional strategy. According to the definition given in section 1.2.4, attainability of delta-neutrality is expressed as the percentage of cases when delta-neutrality is attained in at least one point. Figure 1.3.13 shows the relationship between delta-neutrality attainability and the number of days left to options expiration (presented separately for calm and volatile markets). As was the case for the delta-neutral strategy, for the partially directional strategy we discovered statistically significant inverse relationships between delta-neutrality attainability and the time remaining to options expiration (for calm period $t = 10.29$, $p < 0.0001$; for volatile period $t = 10.89$, $p < 0.0001$; see Figure 1.3.13).

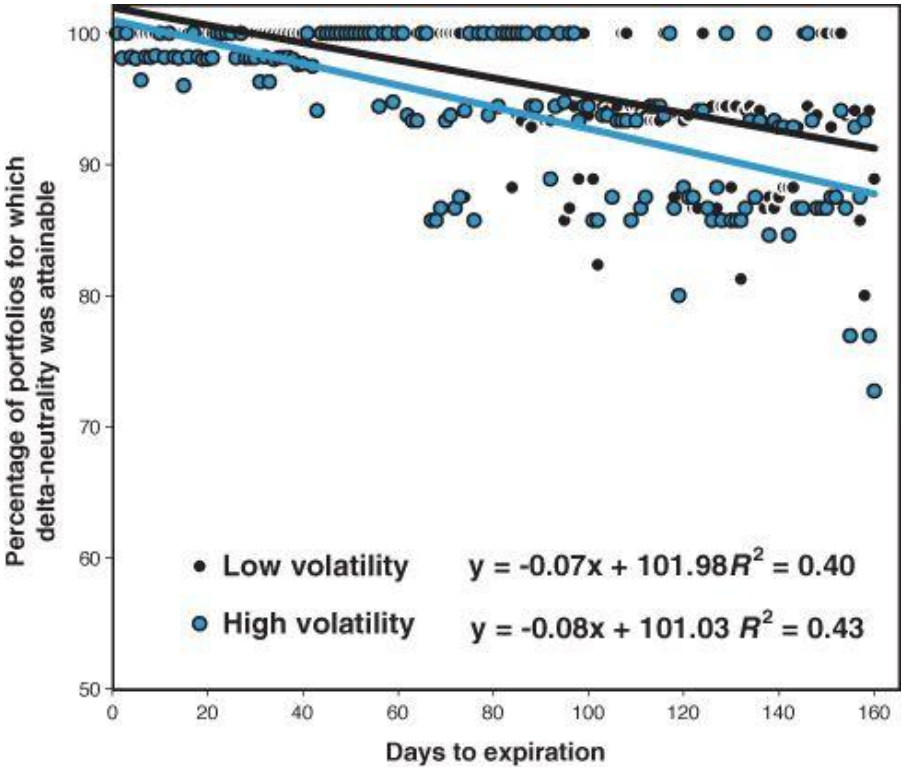

Figure 1.3.13. Relationship between delta-neutrality attainability and the number of days left until options expiration during calm and volatile markets. The data relates to a ten-year historical period.

Delta-neutrality is attainable in almost all instances when portfolios were created less than 40 days prior to the options expiration. If longer-term options are used, delta-neutrality attainability decreases. However, this decrease occurs at a slower rate than was observed in the case of the delta-neutral strategy (compare Figures 1.3.13 and 1.2.10). Market volatility does not exert a statistically significant influence on the relationship between delta-neutrality attainability and the number of days left until expiration. This conclusion follows from the proximity of regression lines relating to calm and volatile periods. The difference in regression lines slopes is statistically insignificant ($t = 1.65$, $p = 0.10$). In this regard the partially directional strategy also differs from the delta-neutral one, for which delta-neutrality attainability does depend on market conditions (attainability is lower under high volatility).

All characteristics of delta-neutrality boundaries relating to the partially directional strategy are summarized in Table 1.3.1. A similar table was presented earlier for the delta-neutral strategy (Table 1.2.1).

Comparison of these tables reveals the differences between these strategies. Next we contrast the two strategies in regard to the effect of different variants of {market volatility × time to expiration} values on characteristics of delta-neutrality boundaries. The influence of more than half of these variants is not the same for the delta-neutral and partially directional strategies. In particular, the following differences should be noted:

		Volatility			
		High		Low	
Time to Expiration	Short	DNB location: DNL location: DNL length: DNA:	*variable criterion threshold* *medium strikes range* *short* *high*	DNB location: DNL location: DNL length: DNA:	*variable criterion threshold* *narrow strikes range* *medium to long* *high*
	Long	DNL location: DNL location: DNL length: DNA:	*variable criterion threshold* *medium to wide strikes range* *short* *medium*	DNL location: DNL location: DNL length: DNA:	*highly variable criterion threshold* *narrow to medium strikes range* *short* *medium*

Table 1.3.1. Dependence of delta-neutrality characteristics (delta-neutrality boundary [DNB] location, DNB length, and delta-neutrality attainability [DNA]) on market volatility and the time from portfolio creation to options expiration.

• *When the market is highly volatile and the portfolio is created using options with the nearest expiration date,* the delta-neutrality boundary of the delta-neutral strategy is situated in the area of low strikes range values, whereas for the partially directional strategy the boundary is shifted toward average values of this parameter. Besides, in the former case the boundary is longer than in the latter case.

• *When the market is in the state of low volatility and the portfolio is constructed of the nearest options,* the delta-neutrality boundary of the delta-neutral strategy is located in the area of low criterion threshold values, whereas for the partially directional strategy it is found along the whole range of this parameter's values. The boundary length is medium in the former case, though in the latter case it may be quite lengthy (from medium to long).

• *In a volatile market, when the portfolio consists of options with the distant expiration date,* the delta-neutrality boundary of the delta-neutral strategy is situated in the area of average values of strikes range, whereas in the case of the partially directional strategy it is located in the area of average-to-

high values of this parameter. In the former case the delta-neutrality is barely attainable, whereas in the latter case its attainability is medium.

• *In a calm market, when the portfolio is created using long-term options,* the delta-neutrality boundary of the delta-neutral strategy is found in the area of low criterion threshold values, whereas for the partially directional strategy this parameter is highly unsteady (fluctuating from low to very high values). Besides, in the former case the delta-neutrality boundary is situated in the area of high strikes ranges, though in the latter case it is shifted toward low and average values. Delta-neutrality attainability is low for the delta-neutral strategy and medium for the partially directional strategy.

1.3.7. Analysis of the Portfolio Structure

In this section we apply the same approaches and calculate the same characteristics as was done in section 1.2.5 to describe the structure of portfolios created by the delta-neutral strategy. As before, we limit values of the main parameters, criterion threshold and strikes range, to the interval spanning from 0% to 25%. All characteristics of the partially directional portfolios will be analyzed in conditions of low and high volatility. Two time intervals (one-week and two-month) from the time of portfolio creation to options expiration will be considered. To compare the structure of the partially directional and the delta-neutral portfolios, we use the data relating to the same dates of portfolio creation and the same expiration dates that were used in section 1.2.6. Since various aspects of portfolio structure have already been described with regard to the delta-neutral strategy, here we limit our discussion to examining the differences between the structure of partially directional portfolios and the delta-neutral ones.

Number of combinations in the portfolio. The relationships between the number of combinations included in the portfolio and the values of the two main parameters, criterion threshold and strikes range, are similar to the patterns observed in the case of the delta-neutral strategy (see Figure 1.2.11). The main difference consists in the fact that for partially directional strategy the quantity of combinations relating to each variant of the {criterion threshold × strikes range} combination is lower. The decrease in the number of combinations indicates the deterioration of portfolio diversification. However, such a decrease may become critical only under conditions of significant restrictions imposed on strategy parameters: high criterion threshold, narrow range of strike prices, and nearest expiration dates. In all other circumstances the number of

combinations would suffice to reach the acceptable diversification level.

Number of underlying assets in the portfolio. This characteristic represents another important indicator of portfolio diversification. During a high-volatility period (regardless of the time left until options expiration), the dependence of the number of underlying assets on the criterion threshold and strikes range does not differ from the relationship observed earlier for the delta-neutral strategy (see Figure 1.2.12). However, in calm market conditions the partially directional strategy demonstrates other forms of relationships (compare Figures 1.2.12 and 1.3.14). While for the delta-neutral strategy (under the condition of using short-term options) the number of underlying assets is independent of the strikes range and decreases exponentially as the criterion threshold increases, in case of the partially directional strategy the widening of the strikes range leads to an increase in the number of underlying assets, and an increase in the criterion threshold causes a slower decrease in the underlyings quantity (see Figure 1.3.14). When long-term options are used, the difference between the two strategies becomes even more obvious. Whereas for the delta-neutral strategy the number of underlying assets is large at low criterion thresholds and almost does not depend on the strikes range, in the case of the partially directional strategy this number turns out to be quite low (regardless of the criterion threshold values) and drops even lower under narrow strikes ranges (see Figure 1.3.14).

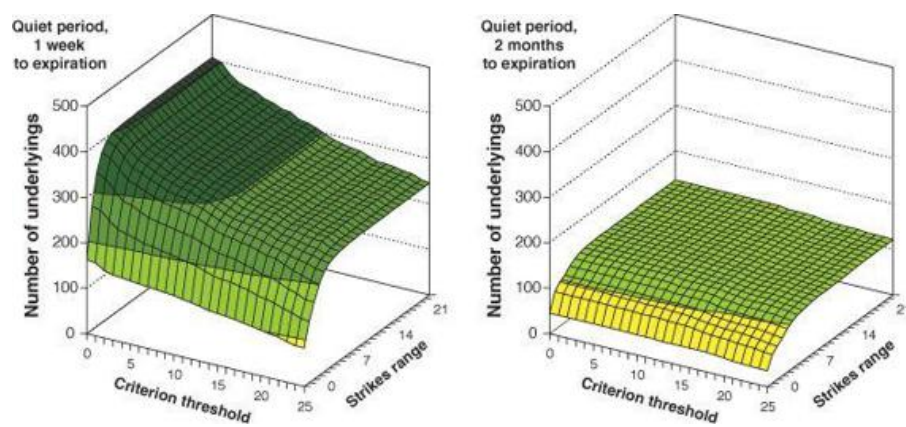

Figure 1.3.14. Dependence of the number of underlying assets on the criterion threshold and strikes range. Two time periods (one-week and two-month) from the time of portfolio creation (in calm market conditions) until options expiration are shown.

Proportions of long and short combinations. During a highly volatile period, all portfolios created within the framework of the

partly directional strategy consist mainly of short combinations. The same was observed for the delta-neutral strategy. In calm market conditions the percentage of short combinations in the portfolio is independent of the width of strikes range, but depends on the criterion threshold value (this also coincides with patterns detected for the delta-neutral strategy; see Figure 1.2.13). To better trace the relationship between the percentage of short combinations and the criterion threshold and to compare two strategies with each other, we averaged the data relating to different strikes ranges (this procedure reduces chart surfaces to lines). Regardless of the time left until options expiration, the percentage of short combinations in the portfolio decreases as the criterion threshold increases. This decrease is going faster and reaches lower percentages of short combinations in portfolios consisting of the nearest options than in portfolios constructed of more distant options. The same trends are peculiar for the delta-neutral strategy (see Figure 1.3.15). The main difference between the two strategies is that when short-term options are used, the percentage of short combinations in partially directional portfolios is always lower than in portfolios created by the delta-neutral strategy (regardless of the criterion threshold value). When long-term options are used, the percentage of short combinations in partially directional portfolios is lower (as compared with portfolios created by the delta-neutral strategy) at low criterion threshold values, but becomes higher if the criterion threshold exceeds 10%.

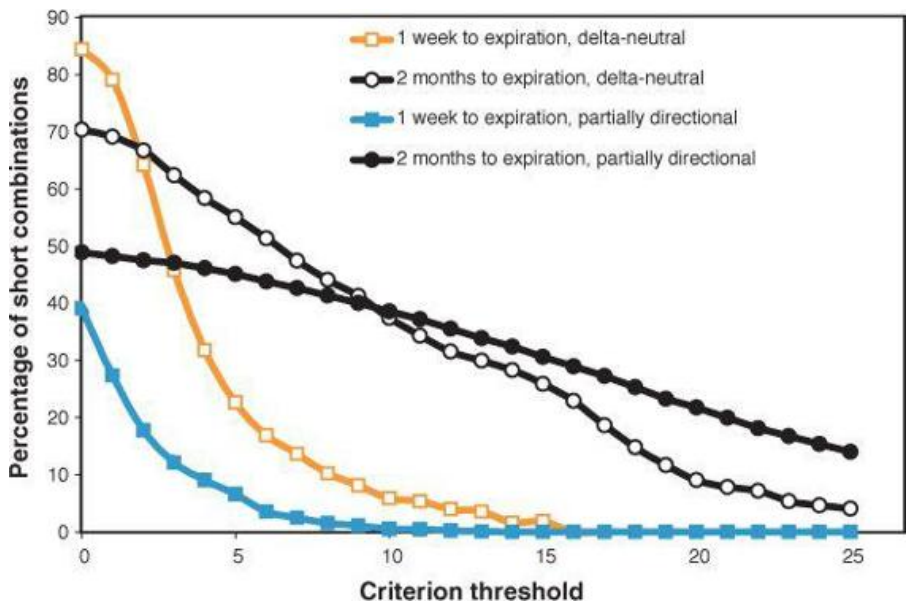

Figure 1.3.15. Relationship between the percentage of short combinations and the criterion threshold. Two strategies and two time periods from the time of portfolio creation until options

expiration are shown.

Portfolio asymmetry. This characteristic expresses the extent of asymmetry of the portfolio payoff function with respect to the current index value. Since the concept underlying the partially directional strategy includes the forecast component and does not require market-neutrality, the asymmetry coefficient may rise to quite high values.

While for the delta-neutral strategy this characteristic did not exceed 0.35 (for all values of criterion threshold, strikes range, market volatility, and time left to options expiration; see Figure 1.2.16), in the case of the partially directional strategy this indicator rises to the range of 0.4 to 1.4. Notwithstanding that, reduction of asymmetry to its minimum (while keeping the forecast element) is an important task in developing the partially directional strategy. Hence, it is necessary to search for such parameter sets that would decrease portfolio asymmetry as much as possible. The dependence of the asymmetry coefficient on the criterion threshold and strikes range values is shown in Figure 1.3.16 for portfolios consisting of short-term options. During a calm market, the asymmetry of the portfolio payoff function reaches its maximum under the weakest restrictions imposed on the strategy parameters (a low criterion threshold and a wide range of strike prices). In a volatile market the maximum asymmetry is observed at higher values of criterion threshold. The important fact is that during both calm and crisis periods there are a large number of {criterion threshold × strikes range} variants for which portfolio asymmetry can be maintained at the acceptably low level.

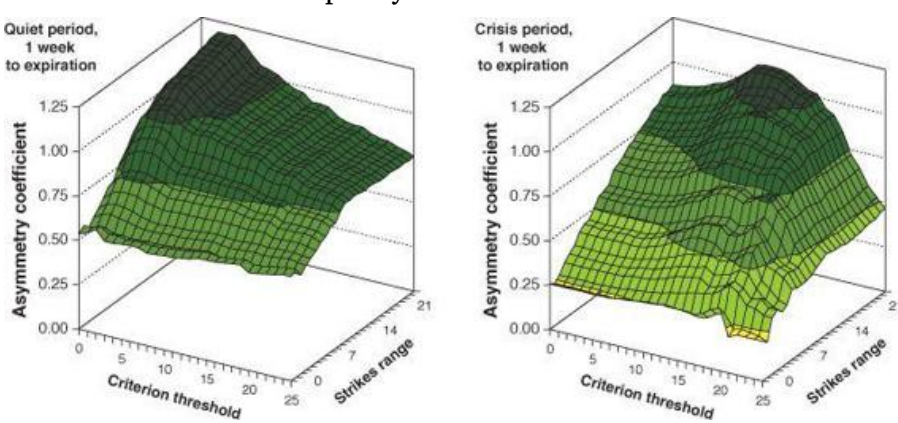

Figure 1.3.16. Dependence of the asymmetry coefficient on the criterion threshold and strikes range. Two market conditions (calm and crisis) are shown for portfolios created using one-week options.

Loss probability. The relationships between the loss probability and the values of the two main parameters, criterion threshold and strikes

range, resemble by their shape the relationships detected for the delta-neutral strategy. This similarity holds true for both volatility levels (see Figure 1.2.17). The only difference (which holds only during a calm period and only for portfolios constructed of long-term options) is that, while for the delta-neutral strategy loss probability increases as the strikes range widens, in the case of partially directional strategy this parameter has no effect on loss probability. For all variants of {strikes range × criterion threshold} the absolute values of loss probability relating to the partly directional portfolios are slightly higher as compared to the estimates of this characteristic for portfolios created by the delta-neutral strategy.

VaR. In the case of the delta-neural strategy we observed that the time left until options expiration influences portfolios VaR more strongly than market volatility does (see Figure 1.2.19). The same phenomenon can be noted for the partially directional strategy (VaR increases several times if two-month options are used instead of one-week contracts). The VaR of portfolios created within the framework of the delta-neutral strategy reached its maximum at the lowest criterion threshold values and the widest strikes ranges. In the case of the partially directional strategy, the peak of VaR values shifts toward higher criterion thresholds; it is especially noticeable during the high-volatility period (see Figure 1.3.17).

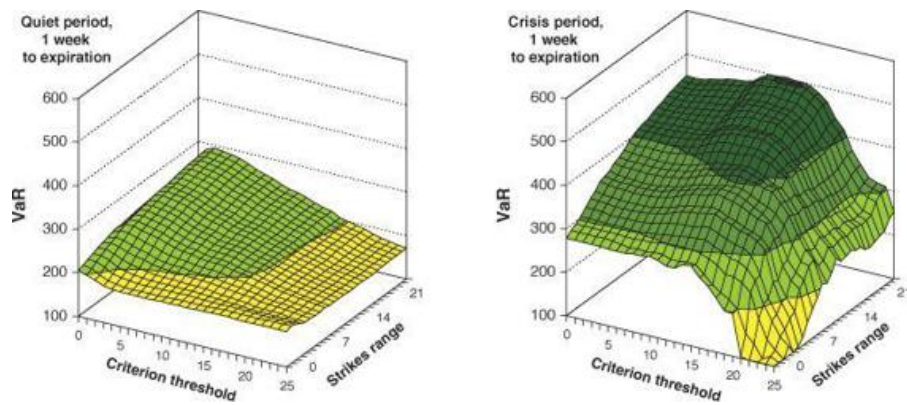

Figure 1.3.17. Dependence of VaR on the criterion threshold and strikes range. Two market conditions (calm and crisis) are shown for portfolios created using one-week options.

1.4. Delta-Neutral Portfolio as a Basis for the Option Trading Strategy

In preceding sections we examined two main aspects of delta-neutral and partially directional option strategies: (1) the possibility to attain delta-neutrality, location, and length of delta-neutrality

boundaries, and (2) the structure and properties of obtainable option portfolios. Here we weave these two aspects into one general concept and develop a technique to select such portfolio that satisfies the requirements of the trading strategy developer as much as possible.

1.4.1. Structure and Properties of Portfolios Situated at Delta-Neutrality Boundaries

Recall that a delta-neutrality boundary represents a set of portfolios with one common quality: Their index delta equals zero. In this respect such portfolios are all identical. However, they differ substantially in many other important characteristics. In fact, these characteristics compose the whole complex of properties that determine expected profitability and potential risks of the trading strategy under construction. In this section we describe the methodology of determining the structure and properties of portfolios situated at the boundaries of delta-neutrality. Although nominally all these portfolios can be used within both delta-neutral and partially directional strategies, the developer should establish the procedure to select a single variant (or several equivalent alternatives) with the best possible combination of characteristics. In this and the next section (devoted to the problem of optimal portfolio selection), we present examples related to delta-neutral trading strategy. However, all methods described here can equally be applied to any option strategy.

The following procedures need to be executed in order to determine the characteristics of delta-neutral portfolios:

1. Present the delta-neutrality boundary as a sequence of points, each of which is specified by a pair of coordinates at the two-dimensional {criterion threshold × strikes range} system of axes. Since in theory the boundary may consist of an infinite number of points, a discrete step has to be defined in order to determine point coordinates. We set this step at 1% for both criterion threshold and strikes range parameters.

2. Present the dependence of the characteristic under investigation on the criterion threshold and strikes range as a topographic map. (Earlier these relationships were presented as three-dimensional charts; see Figures 1.2.11 through 1.2.13 and others.) Horizontal and vertical axes of the map correspond to values of criterion threshold and strikes range parameters, respectively. The altitude of each point on the map expresses the value of a certain characteristic of the portfolio defined by the pair of {criterion threshold × strikes range} coordinates.

3. Plot the delta-neutrality boundary on the topographic

map. Coordinates of points composing the boundary indicate locations of delta-neutral portfolios. Altitude marks corresponding to these locations represent the values of the characteristic under investigation.

4. Reiterating steps 1 to 3 for each characteristic, we obtain the complete set of characteristics for each delta-neutral portfolio.

The results of executing the first three procedures for the "loss probability" characteristic are shown in Figure 1.4.1. In this example we used the same data as in sections 1.2.5 and 1.3.7: The portfolios creation date is January 11, 2010, and the expiration dates are January 15, 2010 (for one-week options), and March 19, 2010 (for two-month options). Points and boundaries of delta-neutrality were determined by applying the method described in section 1.2.2. The topographic map was drawn using the procedures applied earlier to draw Figures 1.2.3 and 1.2.8. Delta-neutrality boundaries and topographic maps, presented in a common coordinate system (see Figure 1.4.1), enable determining the value of the loss probability for each delta-neutral portfolio. Altitude marks of the topographic map show the loss probability for each portfolio located at the delta-neutrality boundary (as well as for all other portfolios not situated at this boundary, which we are not really interested in).

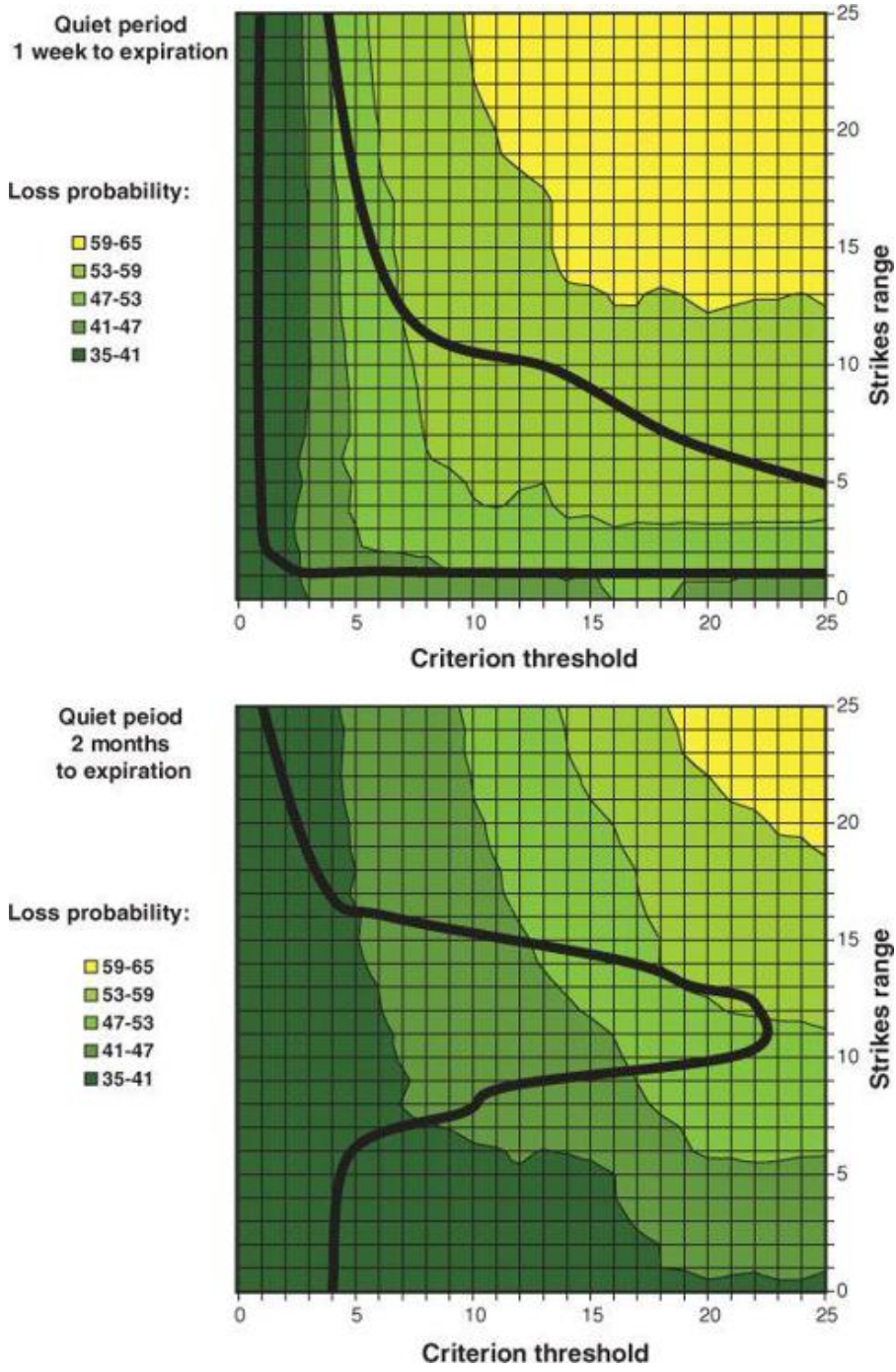

Figure 1.4.1. Plotting delta-neutrality boundaries (bold lines) on the topographic map of the "loss probability" characteristic. Data relate to calm market conditions and to two time periods from the time of portfolio creation until options expiration (one-week and two-month).

Figure 1.2.1 illustrates the procedures of determining the values of a certain characteristic for the complete set of delta-neutral portfolios obtainable under given conditions. However, it does not enable getting the exact values of this characteristic (because altitude marks on the topographic map are presented in the form of interval bands). Furthermore, creation of automated trading strategies cannot be built on visual analysis, but requires developing computational algorithms. Let us demonstrate the procedure of determining the exact values of the "loss probability" characteristic for the data presented on the upper map of Figure 1.4.1 (when portfolios were created using options with the nearest expiration date).

In this particular case two delta-neutrality boundaries were obtained. One of them is at first parallel to the strike price range axis and then turns sharply and goes along the criterion threshold axis (we will not consider this boundary since all portfolios located on it have very low values of at least one of the parameters). The second boundary traverses the topographic map from the top-left corner toward the bottom-right part of the map. Portfolios situated on this boundary have various combinations of {criterion threshold × strikes range} values. To make things clearer, we replace the topographic map with a table in which the cells reflect loss probabilities relevant to all possible variants of parameters combinations {criterion threshold × strike price range}. In Table 1.4.1 all cells related to the delta-neutrality boundary (which is broken here into separate points) are marked with gray. Note that the disposition of gray cells in this table repeats the shape of the delta-neutrality boundary in the top chart of Figure 1.4.1. Values of gray cells represent loss probabilities of corresponding delta-neutral portfolios. This way of determining portfolio characteristics is rather easy, is simple to program, and eliminates a biased judgment which otherwise is inevitable in visual chart analysis.

Table 1.4.1. Loss probabilities. Each cell represents one portfolio with a certain combination of "criterion threshold" and "strikes range" parameters. Delta-neutral portfolios are marked with gray.

Strike range	Criterion threshold																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
25	33	37	43	48	52	54	56	57	58	59	60	61	61	62	62	63	63	62	63	63	63	63	63	63	62
24	33	37	43	48	52	54	55	57	58	59	60	61	61	62	62	63	63	62	63	63	63	63	63	63	62
23	33	37	42	48	52	54	55	57	58	59	60	60	61	62	62	62	62	62	62	63	63	63	62	62	62
22	33	36	42	48	52	54	55	57	58	59	60	60	61	61	62	62	62	62	62	63	63	63	62	62	62
21	33	36	42	47	51	53	55	56	58	59	59	60	60	61	62	62	62	62	62	62	62	62	62	62	62
20	33	36	42	47	51	53	55	56	57	58	59	60	60	61	61	62	61	61	62	62	62	62	62	62	61
19	33	36	42	47	51	53	55	56	57	58	59	59	60	61	61	62	61	61	62	62	62	62	62	62	61
18	33	36	42	47	51	53	54	55	56	58	58	59	59	60	60	61	61	61	61	61	61	62	61	61	61
17	33	36	41	46	50	52	54	55	56	57	58	58	59	60	60	60	60	60	61	61	61	61	60	60	60
16	33	36	41	46	50	52	54	55	56	57	58	58	59	60	60	60	60	60	61	61	61	61	60	60	60
15	33	36	41	46	50	52	54	55	56	57	57	58	59	60	60	60	60	60	61	61	61	61	60	60	60
14	33	36	41	46	50	52	53	55	56	57	57	58	58	59	60	60	60	60	60	60	60	60	60	60	60
13	33	36	41	46	49	52	53	55	55	56	56	57	58	59	59	59	59	59	59	59	59	59	59	59	59
12	34	36	41	46	49	51	53	54	55	56	56	57	57	58	58	59	59	58	59	59	59	59	59	59	59
11	34	36	41	45	49	51	53	54	55	56	56	56	57	58	58	58	58	58	58	58	58	58	58	58	58
10	34	36	41	45	48	51	52	54	54	55	55	56	56	57	57	58	58	57	58	58	58	57	57	57	57
9	35	37	41	45	48	50	52	54	54	55	55	56	56	57	57	57	57	57	57	57	57	57	57	57	57
8	35	37	41	45	48	50	52	53	54	55	55	56	56	57	57	57	57	57	57	57	57	57	57	57	57
7	36	37	42	45	48	50	51	53	54	55	55	55	55	56	55	56	55	55	56	56	55	55	55	56	56
6	37	39	43	46	49	50	51	53	53	55	55	55	54	55	55	56	55	55	56	56	55	55	55	55	56
5	37	38	42	45	47	49	50	52	52	53	54	53	53	54	54	55	54	55	55	55	54	54	54	54	54
4	37	39	43	46	48	49	50	51	52	53	53	53	52	54	54	55	54	55	55	55	54	54	54	54	54
3	38	40	43	46	47	48	49	50	50	50	51	51	51	52	52	53	53	53	52	53	53	53	53	53	52
2	38	39	42	45	46	47	47	47	48	48	48	49	49	50	49	50	50	50	49	49	49	48	48	48	49
1	38	39	42	45	46	46	46	46	47	47	47	47	47	48	47	48	48	48	47	47	47	47	47	47	47

1.4.2. Selection of an Optimal Delta-Neutral Portfolio

In the preceding section we described the method of finding a set of delta-neutral portfolios and established a procedure to determine their characteristics. The next step is to select out of the whole set of attainable delta-neutral portfolios such a variant that would have characteristics satisfying the requirements of the trading strategy developer in the best way. However, the full match between the requirements and the values of all characteristics is never attainable within one single portfolio. Therefore, the problem of portfolio selection should be formulated in the following way: *Select an alternative possessing a set of characteristics that is closest to some reference standard determined by the strategy developer.*

To demonstrate the main approaches to solving the selection problem, let us consider the complete set of characteristics of delta-neutral portfolios obtained in the two following cases. In one case portfolios are created using one-week options; in another case, using two-month options (both, during a calm market period). Earlier we determined the values of one characteristic, loss probability (the top chart in Figure 1.4.1 refers to the first case; the bottom chart, to the second case). To examine other characteristics, it is necessary to construct a table, similar to Table 1.4.1, for each characteristic and to

fix values of cells related to delta-neutrality points (that is, cells marked with gray). It is important to note that delta-neutrality boundaries of portfolios constructed of two-month and one-week options differ significantly (see Figure 1.4.1). Consequently, the distributions of gray cells (that mark delta-neutrality points) within tables relating to the nearest and to more distant options also differ.

To avoid the overburdening of this book with interim tables, we do not show here separate tables for each characteristic. Instead, we summarize in Table 1.4.2 the characteristics relating only to delta-neutral portfolios (that is, only cells marked with gray). This provides a general overview of a complete set of characteristics relevant to all available variants of delta-neutral portfolios.

Table 1.4.2. Characteristics of delta-neutral portfolios created during a calm market using one-week options. Each line presents a separate portfolio specified by a unique combination of two parameters (criterion threshold and strikes range). Gray marks denote cells that contain values belonging to the interval of the desired values of a specific characteristic. Bold frames indicate three portfolios satisfying requirements for five out of six characteristics.

Parameters		Characteristics					
Criterion Threshold	Strikes Range	Quantity of Combinations	Quantity of Underlyings	% of Short Combinations	Loss Probability	Asymmetry Coefficient	VaR
4	25	459	85	31	48	0.10	220
4	24	458	85	31	48	0.10	220
4	23	454	85	31	48	0.10	220
4	22	452	85	31	48	0.09	221
4	21	447	85	32	47	0.09	220
4	20	441	85	32	47	0.09	210
5	19	341	66	22	51	0.09	212
5	18	331	66	22	51	0.08	212
5	17	321	66	23	50	0.08	215
5	16	321	66	23	50	0.08	215
6	15	264	57	17	52	0.08	214
6	14	259	57	17	52	0.08	212
6	13	243	56	16	52	0.07	216
7	12	200	51	13	53	0.07	213
8	11	176	47	9	54	0.06	216
9	11	194	44	8	55	0.05	214
10	10	134	40	6	55	0.05	214
11	10	122	37	6	55	0.05	211
12	10	110	32	4	56	0.05	215
13	10	104	31	3	56	0.05	217
14	10	92	26	1	57	0.05	213
15	9	79	21	1	57	0.05	206
16	8	70	19	0	57	0.04	211
17	8	68	18	0	57	0.05	212
18	7	56	18	0	55	0.04	220
19	7	51	16	0	56	0.04	214
20	6	47	16	0	56	0.04	207
21	6	46	16	0	55	0.04	209
22	6	41	15	0	55	0.04	206
23	5	37	13	0	54	0.04	216
24	5	36	13	0	54	0.04	215
25	5	33	12	0	54	0.04	214

The values of six characteristics are shown in Table 1.4.2 (portfolios consisting of short-term options) and Table 1.4.3 (portfolios consisting of long-term options) for all obtainable delta-neutral portfolios. To analyze these data and to perform the selection of the optimal portfolio on the basis of this analysis, it is necessary to fix the intervals

of the desired values for each characteristic. These intervals depend on the peculiarities of the strategy under development, as well as on individual preferences of the developer and on external restrictions imposed on him. Thus, in each specific case the intervals of the desired values may be different. We will use the following intervals:

- The number of combinations: from 20 to 200
- The number of underlying assets: from 20 to 100
- The percentage of short combinations: from 20% to 80%
- Loss probability: less than 50%
- Asymmetry coefficient: less than 0.1
- VaR: less than 600

Table 1.4.3. Characteristics of delta-neutral portfolios created during a calm market using two-month options. Each line presents a separate portfolio specified by a unique combination of two parameters (criterion threshold and strikes range). Gray marks denote cells that contain values belonging to the interval of the desired values of a specific characteristic. A bold frame indicates the single portfolio satisfying requirements for five out of six characteristics.

Parameters		Characteristics					
Criterion Threshold	Strikes Range	Quantity of Combinations	Quantity of Underlyings	% of Short Combinations	Loss Probability	Asymmetry Coefficient	VaR
4	0	1516	441	70	40	0.08	837
4	1	1362	411	69	40	0.08	815
4	2	1199	362	68	39	0.08	800
4	3	992	313	65	39	0.08	800
4	4	858	275	62	39	0.08	796
4	5	737	242	60	39	0.08	791
5	6	775	219	54	40	0.10	780
7	7	1028	191	48	41	0.13	761
10	8	1423	179	38	42	0.18	718
14	9	1884	167	39	44	0.19	654
20	10	2278	158	35	46	0.18	578
21	11	2073	141	30	48	0.16	561
21	12	1822	126	27	50	0.14	548
18	13	1403	118	26	49	0.12	553
16	14	1094	106	24	49	0.11	558
11	15	586	78	25	46	0.10	614
7	16	248	52	25	43	0.08	692
4	17	144	36	33	42	0.06	713
3	18	61	28	18	43	0.06	720
3	19	52	23	10	45	0.04	715
3	20	47	19	6	45	0.04	718
2	21	25	12	9	43	0.03	731
2	22	23	11	9	43	0.04	728
2	23	22	9	6	43	0.03	731
1	24	12	8	8	43	0.02	728
1	25	10	7	8	42	0.02	726

In Tables 1.4.2 and 1.4.3 we used gray to denote cells that contain values belonging to the intervals of the desired values. Within each column gray marks indicate cells relating to the desired interval of a given characteristic. Since each line in these tables represents a separate portfolio, gray marks within a given line denote the characteristics that have the desired values for a specific portfolio.

When portfolios were created using the nearest options (see Table 1.4.2), none of them had a set of characteristics entirely satisfying our requirements (with regard to belonging to the intervals of desired values). However, three portfolios prove in line with five characteristics. These portfolios are denoted with bold frames in Table

1.4.2. Whether at least one of these three portfolios is suitable for utilizing within the framework of particular trading system depends on the selection algorithm adopted by the strategy developer (in the following text we dwell on this subject in detail).

Similarly to the preceding example, when portfolios were constructed of long-term options, none of them has a complete set of characteristics satisfying our requirements (see Table 1.4.3). However, one portfolio has five out of six characteristics that meet the requirements of the intervals of acceptable values. This portfolio is denoted with a bold frame in Table 1.4.3. If that is enough according to the algorithm of portfolio selection, this alternative may be used to open trading positions.

We can propose a variety of algorithms for selecting an optimal delta-neutral portfolio. All of them represent, in essence, different variants of solving the problem of multicriteria selection, where each characteristic represents a separate criterion.

The most restrictive algorithm may be as follows. At the first stage we select portfolios with all characteristics satisfying the a priori established requirements. (In the two examples presented in Tables 1.4.2 and 1.4.3, there were no such portfolios.) At the second stage there are several alternative procedures to implement. All characteristics can be ranked by their significance. After that, one portfolio with the best value of the most significant characteristic is selected from the set of portfolios that have been chosen at the first stage. If there is more than one such portfolio, further selection is conducted on the basis of the characteristic that is ranked second in importance, and so on. The drawback of this method consists in the objective difficulty in ranging different characteristics in the order of their relative importance (many of them are equally significant). Another procedure that can be implemented at the second stage is the Pareto method of multicriteria analysis. (Applying this method to options was described by Izraylevich and Tsudikman [2010]). However, in this case we cannot control the number of selected portfolios. This drawback may turn out to be very substantial since the selection of several portfolios instead of a single one requires opening much more positions that may significantly increase losses caused by slippage and operational costs.

A less restrictive algorithm may take the following form. At the first stage we select all portfolios that have n characteristics satisfying the requirements for the interval of desired values. The total number of characteristics is m. In our examples $m = 6$. If n is set at 5, in the example presented in Table 1.4.3 a single portfolio will be selected at the first stage. In Table 1.4.2 there are three such portfolios. The

second stage may be realized in the same way (or ways) as in the more restrictive algorithm. For example, if the developer of the strategy decides that the "number of combinations" is the most significant characteristic (the lower the number of combinations, the better, but not less than 20), then at the second stage the portfolio defined by parameters {criterion threshold = 4, strike price range = 20} will be selected out of three alternatives. Alternatively, the second stage can be realized in the following way. Within the variants that have been chosen at the first stage, we can select at the second stage the portfolio possessing the best value for the failed characteristic. Although this best value does not fall into the interval of the desired values of this particular characteristic, it still dominates over all other variants selected at the first stage. In Table 1.4.2 all three portfolios that have been selected at the first stage have unsatisfactory values for the "number of combinations" characteristic. However, the portfolio defined by parameters {criterion threshold = 4, strike price range = 20} has better values of these two failed characteristics than the other two variants. This portfolio may be selected as the optimal one.

For both more restrictive and less restrictive algorithms the implementation of the second stage may be based on another principle. Instead of a priori ranking of characteristics by their relative significance, we can consider the characteristic that varies in the widest range of values as the most important one. For example, in Table 1.4.2 the values of "asymmetry coefficient" do not change at all. Therefore, all three portfolios selected at the first stage do not differ from each other by this characteristic. Hence, there is no sense in selecting it as the main reference points at the second selection stage. On the other hand, values of the other two characteristics, "number of combinations" and "VaR," fluctuate in a wider range of values. Thus, in this particular case, it would be natural to use them as the main characteristics for the final selection of an optimal delta-neutral portfolio.

It is important to emphasize that whichever algorithm is adopted in the process of developing the automated trading strategy, it represents, in many respects, the main factor determining the choice of the delta-neutral portfolio that will be used to open trading positions.

Chapter 2. Optimization

2.1. General Overview

The problem of selecting the optimal solution arises in all spheres of human life—on individual, corporate, and state levels. The optimization may be defined as the search for either an optimal structure of a certain object (structural optimization) or as a sequence of actions (calendar optimization). In the context of creating automated trading strategies, *parametric optimization* is the subject of interest. In this case the optimal solution represents a combination of parameter values.

Despite the availability of numerous sophisticated methods that have been developed recently for treating various aspects of optimization, it is still impossible to propose a universal approach that would apply equally well to all situations. The reasons include the heterogeneity of optimization problems, limitedness of time, and computational resources. Only a synthetic approach combining achievements from various mathematic fields can provide adequate framework for optimization of trading strategies.

2.1.1. Parametric Optimization

Depending on the definition of the problem, parameters may be *real numbers* (for example, the share of capital invested in a certain strategy) or *integers* (number of days from position opening to options expiration), or they may be presented by dummy variables (for example, if the parameter signifies whether we use a certain type of option combination, it can assume values of 1 or 0). The number of parameters may be limited to one (one-dimensional optimization), but in most cases there is more than one of them (multidimensional optimization).

The definition of an optimization problem may be unconditional or contain certain constraints. In particular, not all possible combinations of parameter values are permissible. Due to specific limitations, some of them may be unacceptable or unrealizable. If some combinations of parameter values are excluded from consideration, the optimization is defined as conditional. The constraints may look like

$$x_1 + x_2 + \ldots + x_n = M,$$

where x_i takes on a value of 0 or 1, depending on whether the trading position is opened for underlying asset i. The point of this constraint may be, for example, that the total number of underlying

assets in the portfolio equals precisely M. Constraints may also be defined as inequality:

$$c_1x_1 + c_2x_2 + \ldots + c_nx_n \leq K.$$

Here c_i may be the price of option i and x_i is the number of options i in the portfolio. The sign indicates whether the position is long (plus) or short (minus). The meaning of the constraint is that the total value of the option portfolio does not exceed K.

Constraints may also be applied to the range of values that a parameter may possess (in the previous chapter we often used the notion of "the range of acceptable values"). Such constraints are often applied in the development of automated trading strategies. They may be applied on the basis of practical reasons, since the reduction of the set of acceptable values reduces substantially the volume of computational work and shortens the optimization time. Besides, constraints may follow from some specific features of the strategy under development or from requirements of the risk management system. At last, constraints on the range of acceptable values may arise when the data for calculation of an objective function are unavailable for certain parameter values or such calculation is impossible.

In order to avoid a mess in definitions used in the description of optimization procedures, we provide a short explanation for the terms that we use in this chapter.

- *Optimization space*—A set of all possible combinations of parameter values.
- *Node*—The smallest structural element of an optimization space (a unique combination of parameter values). A full set of nodes forms an optimization space.
- *Computation*—The whole set of procedures necessary to calculate the objective function for one node of an optimization space.
- *Full optimization cycle*—The whole set of all computations performed in the course of searching for an optimal solution (from the beginning of the optimization procedure until the algorithm stops).
- *Objective function*—The quantitative indicator expressing the extent of utility of a given combination of parameter values from the point of view of a trading system developer (may be calculated either analytically or algorithmically).
- *Global maximum*—The node with the highest value of objective function. The global maximum can be viewed as

the highest peak of the two-dimensional optimization space. There may be more than one global maximum.

• *Local maximum*—The node situated at one of the peaks of the optimization space but with a lower objective function value than the global maximum. There may be several local maximums.

• *Optimal solution*—The node at which the optimization algorithm stops. This node is specified by a set of the parameter values and by the value of the objective function. Optimal solution does not necessarily coincide with the global maximum. The higher the effectiveness of the optimization algorithm, the closer the optimal solution is to the global maximum.

• *Robustness of optimal solution*—The extent of the objective function variability in the area of optimization space that surrounds the optimal solution node. Although the term "robustness" is widely used in statistics, economics, and biology, it has no strict mathematical formalization when applied to optimization issues. We define the solution as robust if the values of the objective function of most nodes around it are not significantly lower than the objective function of the optimal solution node.

• *Optimal area* is the area of the optimization space where all nodes have relatively high values of the objective function (higher than a predefined threshold). There may be several optimal areas. Usually these areas are situated around nodes of global and/or local maximums. However, the optimal area may also appear as an elevated smooth plateau without evident extremes (in case of the two-dimensional optimization space).

2.1.2. Optimization Space

Optimization space is a set of combinations of all possible parameter values (a full set of nodes). It is defined by three characteristics:

1. Dimensionality, which is the number of parameters to be optimized

2. Range of acceptable values for each parameter

3. Optimization step

Most optimization problems, which are a subject of practical interest for developing an automated trading system, are

multidimensional. This means that the objective function is calculated on the basis of many parameters (it has more than one argument). In complex trading strategies, the number of optimized parameters may be very high. In these cases the risk of overfitting increases greatly (we will dwell on this issue in Chapter 5, "Backtesting of Option Trading Strategies"). Since most objective functions used in optimization of trading systems cannot be calculated analytically, the methods of direct optimization (when the objective function is calculated algorithmically) have to be applied. The complexity of the algorithm depends directly on the dimensionality of the objective function (that is, on the number of its arguments-parameters).

The range of acceptable values is determined by constraints (established by the trading system developer) with respect to the parameters used to calculate the objective function. For example, in Chapter 1, "Development of Trading Strategies," we considered two parameters: a criterion threshold and a range of strike prices. The range of acceptable values for the first parameter was from zero to infinity. The logic behind this choice was the following. Since we used expected profit as the valuation criterion, it was quite natural not to consider any negative range of expected profit values. For the "strikes range" parameter the range was set to be from 0% to 50%. The lower limit was based on the fact that this parameter cannot be negative. The reason for the upper limit was that it is impractical to use strikes that are situated too far from the current price of the underlying asset (distant strikes have low liquidity and wide spreads).

The best method of direct optimization requires calculation of the objective function for all acceptable values of the parameter (exhaustive search). However, in practice this method is rarely implemented since the number of acceptable values may be too high. If the parameter is of integer nature, the number of its values is finite (within the acceptable range not including infinity). Nevertheless, even in this case the exhaustive search through all values may require too many computations and too much time. If the parameter is continuous, the number of its values is infinite without any regard to the range of acceptable values. In this case we have to set a step for its value change (referred to as the "optimization step") and investigate the parameter by moving along its range of acceptable values discretely. The step represents a distance between two adjacent nodes. The higher the step value, the less nodes will be tested and, consequently, the lesser time will be required to complete the full optimization cycle. However, if the step is too wide, the risk of missing the global maximum increases (a sharp peak may get into the interval between two nodes).

The shape of the optimization space has a direct impact on the effectiveness of the optimization procedure and its final outcome. If optimization is one-dimensional (when there is only one parameter), the optimization space may be viewed in a system of axes as a line with coordinates corresponding to the parameter values (axis *x*) and the objective function value (axis *y*). If this line has only one global maximum, the objective function (and optimization space) is *unimodal*. If the objective function has one or more local maximums apart from the global one, it is called *polymodal*. If the objective function has approximately equal values at the whole range of parameter values, it is nonmodal and can hardly be used for optimization of this particular parameter.

In the case of two-dimensional optimization (when objective function is calculated on the basis of two parameters), the optimization space can be easily presented as a surface. Such a surface is depicted as a topographic map with axes corresponding to parameters and altitude marks—to values of the objective function. Unimodal surfaces have a single peak, while polymodal ones have a multitude of such tops. A relatively flat surface is nonmodal and can hardly be appropriate for optimization.

In the three-dimensional case (and dimensionality of higher orders), nodes represent areas in which values of all parameters are relatively high. In these cases it would be impossible to depict the optimization space topographically, but this is not really necessary since computational algorithms do not need our imagination.

Most optimization methods perform better when searching within the simplest optimization space—unimodal space with a single maximum. If there are several local maximums and the search is not exhaustive, most methods produce a sub-optimal solution, which can be more or less far from the best one.

Optimization space has a number of properties that influence optimization effectiveness significantly. Two of them are worth mentioning. The first property is *optimization space smoothness*. In the two-dimensional case, smoothness means absence of a large quantity of small local peaks making the surface “hilly.” In extreme cases, optimization space may be either absolutely smooth (with a single extreme) or totally bumpy with a lot of sharp peaks and bottoms. Obviously, smooth space is preferable from the optimization efficiency standpoint. Hilly space increases the risk of stopping the optimization procedure at a small local extreme. Later we will show (section 2.7.2) that the smoother the space, the higher the effectiveness of different optimization methods and the higher the likelihood of finding an appropriate optimal solution.

The second important property is the *steadiness of optimization space*. Steadiness can be viewed as nonsensitivity of the space relief to (or, in other words, invariability of the space shape depending on) small changes in parameters, which are not involved in the optimization process, but are fixed based on particular reasons. The nonsensitivity to small changes in the strategy structure also belongs to this property. Another aspect of steadiness is the extent of optimization space sensitivity to the length of historical price data used to calculate objective function values. A very short price history results in developing a trading strategy that is adapted merely to recent market trends. On the other hand, a very long price history may lead to adapting the system to possibly outdated data. Besides, it is highly desirable that historical data should reflect different market conditions (calm and crisis periods). In order to optimize a trading system properly, a developer must take these arguments into account.

2.1.3. Objective Function

The essence of parametric optimization consists of finding the highest or the lowest value of the objective function. This function may be specified as an equation, may be specified as a computational algorithm (which calculates the function's values on the basis of a given set of parameters), or may be obtained experimentally. The choice of any particular method to be used in the automated system for finding the optimal solution depends largely on the properties of the objective function.

Following established scientific traditions, optimization problems are solved by finding the lowest value of the objective function. Although finding maximum values and finding minimum values represent opposite problems, the same methods are used to perform both of these tasks. Notwithstanding the traditional approach, we will formulate optimization problems as a search for maximums. (An alternative approach is to find the reciprocal values for the objective function and then perform the optimization by solving the minimization problem.) The reason is that one of the main objective functions used in the optimization of trading strategies is profit and its various derivatives. From a psychological point of view, it is more comfortable to maximize profit than to minimize it.

Historically, optimization theory operated mostly with the objective function given by the analytical formula. In simple cases the formula represents a differentiable function. Its properties (areas of increase and decrease, extreme points) are investigated using derivatives. Equating derivatives to zero in all parameters and solving the obtained system of equations is an elegant way for finding the optimal

solution.

Modern needs to optimize complex heterogeneous systems, supported by impressive theoretical achievements, led to a vast expansion of the range of solvable problems. In most of them the objective function is not given analytically and hence cannot be analyzed with the help of derivatives. In these cases function values are obtained via algorithmic calculations.

Methods that are based on algorithmic calculations and do not require obtaining the objective function derivatives are called direct methods. The obvious advantage of direct methods is that they do not require the objective function to be differentiable. Moreover, it may be not given analytically (in the form of an algebraic equation). The sole requirement to direct methods consists in the necessity to estimate objective function values. Almost all problems involved in the optimization of trading systems are solved on the basis of algorithmic models. Hence, in this chapter we deal exclusively with these methods.

Solving an optimization problem becomes a significantly harder task when there is more than just one objective function. This issue arises frequently in optimization of trading strategies. Not surprisingly, the main objective function for most strategies is profit. However, it would be unwise to limit the whole system to this single characteristic. It is also necessary to consider profit variability, maximum drawdown, share of profitable transactions, and many other important factors, each of which represents a separate objective function.

The specific feature of using several objective functions is that the maximum of one function rarely coincides with the maximum of another one. On the contrary, different objective functions usually contradict each other—optimal values of one of them may be way too far from the maximum of another function. (A similar situation is discussed in section 1.4.) Developing techniques for the effective application of several objective functions is the subject of multicriteria optimization theory. Three main approaches to multicriteria optimization should be mentioned.

1. *Choosing a certain criterion (objective function) as the main one* and considering other criteria as constraints or filters. After obtaining the optimal solution according to the main criterion, other criteria values are calculated at the same node of optimization space. If the solution, derived on the basis of the main criterion, meets constraints set for the secondary criteria, the algorithm stops. If values of these

criteria appear unacceptably low or high, this solution is rejected and the search by the main criterion recommences.

2. *Constructing a combined criterion (convolution)*. It may be calculated as a simple or a weighted arithmetical average (weights reflect the importance of the criteria) or as a geometrical average (also simple or weighted). Besides, there are several other convolution methods.

3. *Using the Pareto optimization method*. In most cases this method leads to finding several optimal solutions even if each criterion has a unique maximum. The outcome of the optimization is a Pareto set that represents a group of solutions that dominate over all other variants not included in it. None of the solutions belonging to the Pareto set dominates over other solutions included in it. Thus, any solution belonging to the Pareto set can be selected as optimal.

All methods of multicriteria analysis introduce a subjective element to the optimization system. In the first method, the subjectiveness consists in selecting one criterion as the main one and defining specific constraints for secondary criteria. In the convolution approach, selecting a particular technique to combine criteria and setting weights (especially if they are introduced to account for the relative importance of different criteria) is highly subjective. The necessity to select the unique solution from the set of equivalent ones (Pareto method) may also be affected by subjective factors. In section 2.4 we discuss the main aspects of multicriteria optimization as applied to the development of automated trading systems.

2.2. Optimization Space of the Delta-Neutral Strategy

The shape and properties of the optimization space depend on many factors, most of which were described in the preceding section. Certainly, the shape of the optimization space is specific for different option strategies. Each strategy has its own set of parameters, their acceptable value ranges, and optimization steps. Therefore, it is quite natural that different strategies have different optimization spaces. Furthermore, even in cases when parameters, acceptable values ranges, and optimization steps are the same for two different strategies (for example, this situation can be encountered for delta-neutral and partially directional strategies), their optimization spaces may still differ substantially (and in most cases they do differ).

In this chapter we will examine the shape and the properties of a

typical optimization space, taking basic delta-neutral strategy as an example. Two main parameters of this strategy (covered in Chapter 1) are fixed at the following values: criterion threshold > 1%, strikes range of 10%. In the preceding chapter we partly touched on the optimization issue while discussing these parameters. However, in section 1.4 we were detecting optimal values mostly relying on scientific approach and not using specific optimization techniques, which are the topic of this chapter.

We will analyze optimization space corresponding to two parameters of a basic delta-neutral strategy: the number of days to options expiration and the length of historical period used to calculate historical volatility (we will call it "HV estimation period"). The meaning of the first parameter is discussed in Chapter 1. The value of this parameter influences directly the structure of the portfolio. The second parameter reflects the length of the price series employed in HV calculation. Historical volatility, in its turn, is used to estimate values of expected profit, which serves as the source for generation of positions opening signals. Despite the fact that the impact of this parameter is indirect, it also represents one of the most important factors in this strategy, since its values define, to a certain extent, which option combinations will finally be included in the portfolio.

2.2.1. Dimensionality of Optimization

One of the main factors determining the shape of the optimization space is a set of parameters. When a trading system developer formalizes the strategy, the first and one of the main issues is the quantity of parameters requiring optimization. In general, it is advisable to stick to the rule of minimizing the number of parameters. There are two reasons for this. Firstly, the higher the number of parameters, the more degrees of freedom are inherent in the system and, hence, the higher the risk of overfitting. Secondly, higher optimization dimensionality requires excessive calculations, which may not be technically feasible. On the other hand, an unreasonable reduction in the number of optimized parameters may lead to overlooking a potentially profitable strategy. Putting all these pros and cons together, not fewer than two and no more than five parameters are usually used in creating automated trading systems.

One-Dimensional Optimization

One-dimensional optimization is straightforward and the simplest one (although it is rarely used in practice). It can be considered as a special case of a more complex multidimensional optimization. Considering the one-dimensional variant can facilitate the understanding of the difficulties that arise in analyses of complex

optimization spaces. Algorithms utilized for solving multidimensional problems are often reduced to consecutive reiterated executions of one-dimensional procedures.

Figure 2.2.1 shows three examples of one-dimensional optimization space. This chart depicts optimization of the "HV estimation period" parameter of the basic delta-neutral strategy for three fixed values of the second parameter that express the number of days left to options expiration. The range of permissible values for the parameter under optimization is from 5 to 300 days; the optimization step is 5 days. The complete optimization space in this case consists of 60 nodes. This optimization was performed with 10-year historical data. Average profit value is used as the objective function. None of the three optimization spaces presented in Figure 2.2.1 is smooth. This is not surprising since this figure is based on real market data, while only spaces based on analytically given equations may be absolutely smooth. Nevertheless, unavoidable statistical "noise" does not prevent us from distinguishing the main patterns peculiar to each of the lines and classifying these optimization spaces according to their modality.

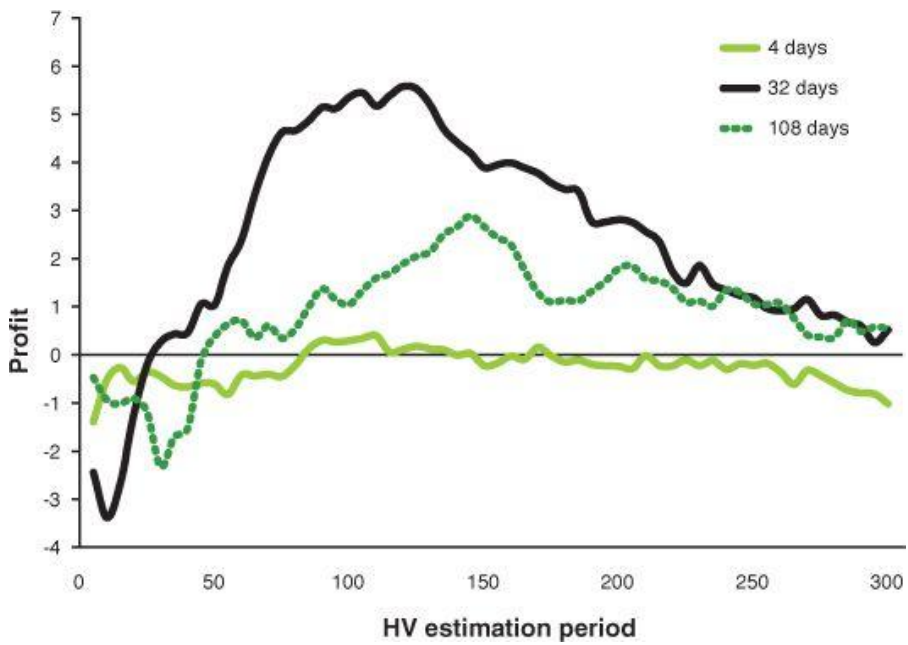

Figure 2.2.1. Three examples of one-dimensional optimization space. Optimized parameter—"HV estimation period." Each line corresponds to one of three values of the fixed (not optimized) parameter "number of days left to options expiration." Objective function: profit.

Each line in Figure 2.2.1 illustrates one of the main forms of optimization space. When the "number of days to options expiration"

parameter was fixed at 32 days, optimization space turned out to be unimodal. In this case, the objective function has the only global maximum corresponding to “120 days” (the length of the historic period used for HV calculation). There were no apparent local maximums. We should note that the statement about the absence of local maximums and unimodality of this optimization space is to some extent a subjective opinion. Actually, since this line is not absolutely smooth, one can claim that some local maximums are present at nodes “105 days” (to the left of the global maximum) and “230 days” (to the right of it). Nevertheless, since on the scale of the whole optimization space these peaks are negligibly small, we consider them as noise. If visual analysis is not obvious, modality determination procedure should be formalized in order to avoid subjective judgments.

The example of polymodal optimization space is presented by the line corresponding to “108 days” as the value of the fixed parameter (the number of days to options expiration). Global maximum of this optimization falls on the “HV estimation period” value of “145 days.” In contrast to the preceding example, the objective function of this optimization has an evident local maximum at the “HV estimation period” value of “205 days.”

Finally, the third line in Figure 2.2.1 is an example of nonmodal optimization space. In this case, when trading positions opening signals were generated only for short-term options (the “number of days to options expiration” parameter is fixed at “4 days”), the objective function fluctuated around zero almost at the whole range of the “HV estimation period” parameter values.

Since in all three examples presented in Figure 2.2.1 the objective functions were analyzed at the whole range of acceptable parameter values (this method is referred to as “exhaustive search”), the selection of optimal solution at first glance is evident. For unimodal objective function it falls on 120 days; for polymodal function, on 145 days; and for nonmodal there is no optimal solution. However, the optimum does not necessarily coincide with the global extreme. There is another (and not less important) criterion for selecting an optimal solution: its robustness. Taking the concept of robustness into account, the selection of 120 days as the optimal solution may not be the best one. Increasing the parameter leads to a sharp decrease in the value of the objective function (which means that this solution is not robust). At the same time, if we select 100 days as the optimal solution, all the adjacent parameter values have sufficiently high values of objective function (which means that this solution is robust). Global maximum of the unimodal objective function represents an example of a more robust solution (since it is situated at the wide plateau) than the

global maximum of the polymodal function (which is located at a rather narrow peak).

Two-Dimensional Optimization

Hereafter we will analyze examples pertaining to two-dimensional optimization space. The same two parameters that were used in the preceding example will form the optimization space: HV estimation period (this parameter was optimized in the preceding example) and the number of days left to options expiration (in the preceding example this parameter was fixed). The range of acceptable values for the first parameter is from 5 to 300 days; the optimization step is 5 days. For the second parameter the range of acceptable values is from 2 to 120 days; the optimization step is 2 days. Thus, the full optimization space consists of 3,600 nodes (60 × 60).

Figure 2.2.2 shows two-dimensional optimization space of the objective function expressing an average profit. In fact, the three one-dimensional spaces discussed earlier (in Figure 2.2.1) represent three special cases of this two-dimensional space. Since Figure 2.2.2 is a topographic surface, vertical sections executed through values 4, 32, and 108 of the "number of days left to options expiration" parameter coincide with one-dimensional profiles shown in Figure 2.2.1. Undoubtedly, the two-dimensional optimization space gives a much broader look at the objective function and better reflects the properties of the trading strategy.

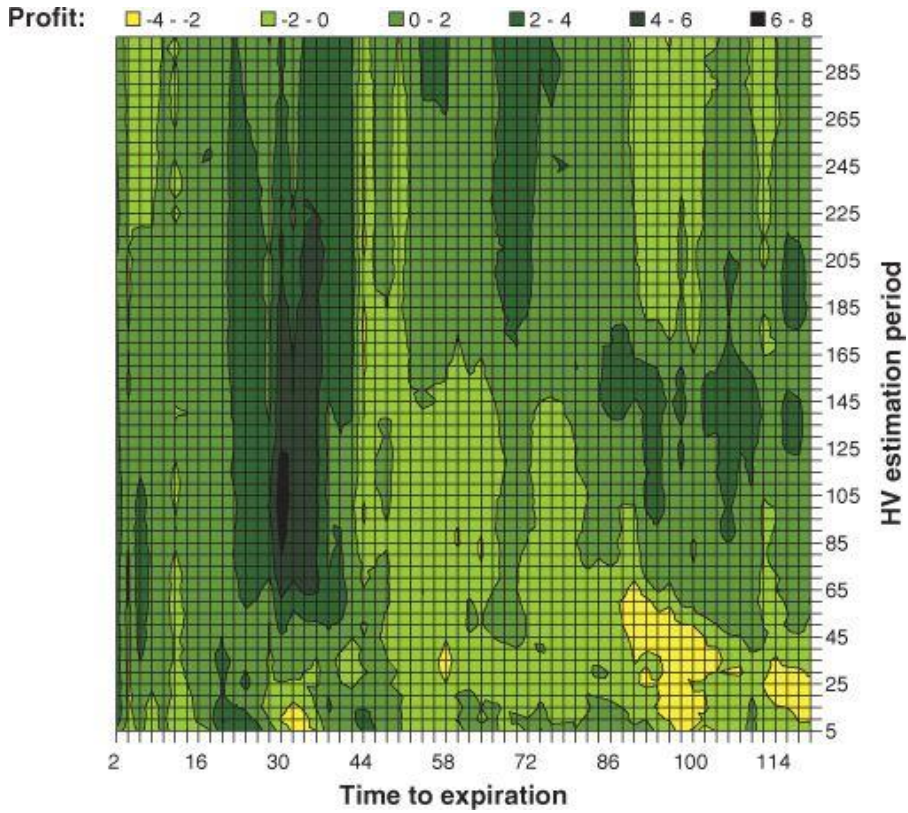

Figure 2.2.2. Two-dimensional optimization space constructed for the basic delta-neutral strategy. Objective function: profit.

The global maximum of the optimization space shown in Figure 2.2.2 has the following coordinates: 30 on the horizontal axis and 105 on the vertical axis. This means that the average profit (which is the objective function) reaches its maximum when positions are opened using options with 30 days left to expiration, and historical volatility used to calculate the criterion is estimated at the historical period of 105 days. This global maximum is situated at the top of a slight edge spreading along the 30th vertical line (the "number of days to expiration" parameter) in the range from 80 to 125 (the "HV estimation period" parameter).

This edge may be considered the optimal area since all nodes inside it have high values of the objective function (>6%). The optimal area itself is also surrounded by a rather wide area consisting of nodes with relatively high objective function values. Therefore, the global maximum of this optimization space is quite robust. However, we should mention that the robustness of this optimal solution is not the same for the two parameters. Changes in the value of the "HV estimation period" parameter inside the optimal area and around it

lead to smaller changes of the objective function than changes in the "number of days to expiration" parameter. Drifting from the optimal value of the latter parameter (30 days) in any direction causes a sharp decrease in the objective function. Hence, the robustness of the global maximum relative to the first parameter is higher than relative to the second one.

The optimization space presented in Figure 2.2.2 is relatively unimodal. This statement is justified by the fact that the optimal area rises distinguishably over the rest of the surface (in the two-dimensional case the optimization space can be called "surface"). At the same time, since this surface is not absolutely smooth, the judgment about its unimodality can be disputed. Apart from the area designated as optimal, this surface has many other areas where the objective function is not just positive, but takes on values falling in a quite good range (2% to 4%). Therefore, depending on the adopted point of view, this surface may, in principle, be regarded as polymodal. Although the issue of classifying the optimization space is not of paramount importance, the presence of local maximums suggests that a global maximum must not necessarily be the best optimal solution. If some local maximum has an objective function value that is not significantly lower than the global maximum, but, at the same time, its robustness is much higher, it may turn out that the best decision would be to select this local maximum as an optimal solution. The objective selection is impossible without applying some quantitative techniques that exclude subjective judgment (this issue is discussed in section 2.5).

To obtain a detailed notion about the shape and properties of the optimization space shown in Figure 2.2.2, we had to calculate the objective function values in all the 3,600 nodes. Since this optimization was performed using 10-year data (like all other optimizations considered in this chapter), the computations relating to one node took about one minute. Accordingly, it took about 60 hours to perform all the calculations for the whole optimization space. Although this was quite reasonable for research purpose, for day-to-day practical work such inflated time costs are not always acceptable, especially if there are more than two parameters. Besides, 3,600 is the number of nodes relating to a single objective function, but usually there are more of them. Hence, in most cases, it would be impractical to make computations for the whole optimization space. Instead, different methods of optimal solution finding that do not require such an exhaustive search should be applied (this issue is discussed in section 2.7).

2.2.2. Acceptable Range of Parameter Values

In this section we discuss how the acceptable range of parameter values influences the optimization space shape. To start the examination of this issue, we cut the acceptable ranges of both parameters by half (relative to ranges used in the preceding section). For the "HV estimation period" parameter the upper limit of the new range will be 150 days (instead of 300), and for the "number of days to options expiration" the limit will be set at 60 days (instead of 120). These constraints lead to four times shrinkage of the original optimization space. This effect of narrowing parameter ranges is positive (since now we need to perform computations for only 900 instead of 3,600 nodes in order to construct the complete optimization space). In case of three-dimensional optimization, cutting the value range by half leads to an optimization space, which is eight times smaller than the original one.

The left chart of Figure 2.2.3 shows optimization space constructed for the new acceptable value ranges. Comparing this smaller space with the original one (in Figure 2.2.2) demonstrates that the area of global maximum was not lost as the result of employing stricter constraints on parameter value ranges. Now this area is situated almost in the center of the space. Besides, the major portion of the original space that contained lower objective function values is now left outside the new optimization space. This means that the share of optimal area relative to the total size of new optimization space has significantly increased. Hence, the probability of finding the global maximum in the process of searching for the optimal solution (using methods not requiring exhaustive search) has also increased. However, when selecting the acceptable value range for a given set of parameters, the developer of the trading system has no idea about the form of the complete optimization space. Consequently, by decreasing the range of acceptable values, the developer may exclude from consideration the area that contains the best solution.

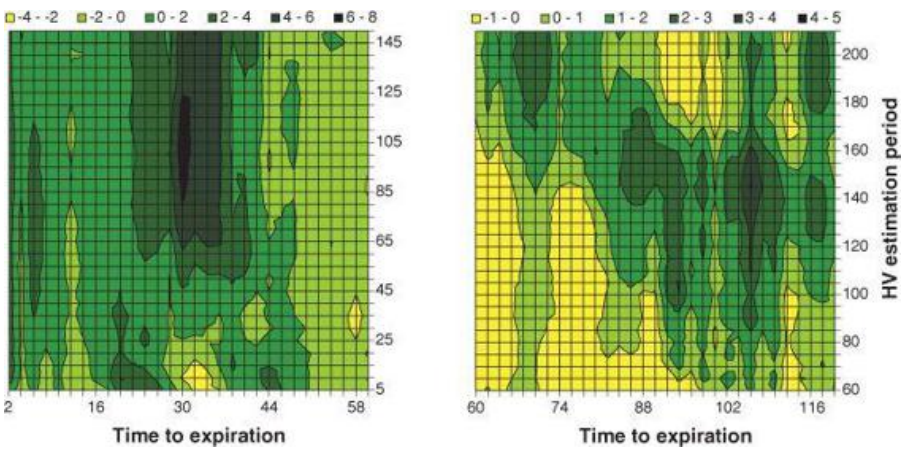

Figure 2.2.3. Optimization spaces of the basic delta-neutral strategy constructed for different acceptable ranges of parameter value. Objective function: profit.

Besides, the range of acceptable values does not necessarily start with the lowest possible value of the parameter (as was the case in the examples given in Figure 2.2.2 and in the left chart of Figure 2.2.3). Suppose that the developer creates a trading strategy employing longer-term options. In this case the developer can fix the lower limit on the range of acceptable values for the "number of days to options expiration" parameter. For example, it can be set at 60 days (let the upper limit be 120). The change in the range of this parameter induces changes in the range of the second parameter, since it seems inappropriate to evaluate long-term options with the criterion calculated on the basis of volatility that has been estimated at the shorter period than the period of options life. Hence, the lower limit on the range of values for the "HV estimation period" parameter should also be set at 60 days (to equate the number of values for both parameters, the upper limit will be fixed at 210 days).

Let us examine the optimization space obtained for these new ranges of acceptable parameter values (the right chart of Figure 2.2.3). Obviously, the results of optimization performed at this new surface will be different. The global maximum falls now on the node with other coordinates—106 on the horizontal axis and 145 on the vertical one. When we analyzed the broader optimization space, this node was just the local maximum. Now when the higher extreme has been left outside the optimization space, the local maximum becomes the global one. The objective function value in this node is 4.1%, which is significantly lower than the global maximum of the original space (7.1%).

To summarize, we can conclude that the range of parameter values determines the shape of optimization space and influences significantly the optimal solution that will finally be chosen. In general, *the broader the range of acceptable parameter values, the higher the chance that the objective function maximum will fall into the optimization space under scrutiny. However, the chance to find this maximum in the process of optimization decreases,* since more nodes have to be examined and the risk that optimization algorithm will stop at a local maximum increases.

2.2.3. Optimization Step

The optimization step does not exert influence on the general shape of the optimization space, but it directly influences its details. The wider the step, the more details of relief of the optimization surface

may be missed in the optimization process. For example, the narrow peak of the objectivity function may be missed because of a wide step. Hence, the scope of information about the objective function decreases when the step is increased.

For the optimization surface analyzed earlier (in Figure 2.2.2) we used the step of 2 days for the "number of days to options expiration" parameter and 5 days for the "HV estimation period" parameter. Now let us increase these values to 4 and 10 days, respectively, and examine the effect of this change on the minuteness of information borne by this new surface. The left chart of Figure 2.2.4 shows the optimization space obtained as the result of increasing the step. Comparison of this surface with Figure 2.2.2 reveals that, despite reduction in details, the area of global maximum remained intact. Originally the global maximum node had coordinates of 30 on the horizontal axis and 105 on the vertical axis. Coordinates of the new global maximum are 30 and 100, respectively. Although the node with the highest value of objective function (7.1%) has disappeared, it was replaced by another global maximum with a very close value of objective function (7%).

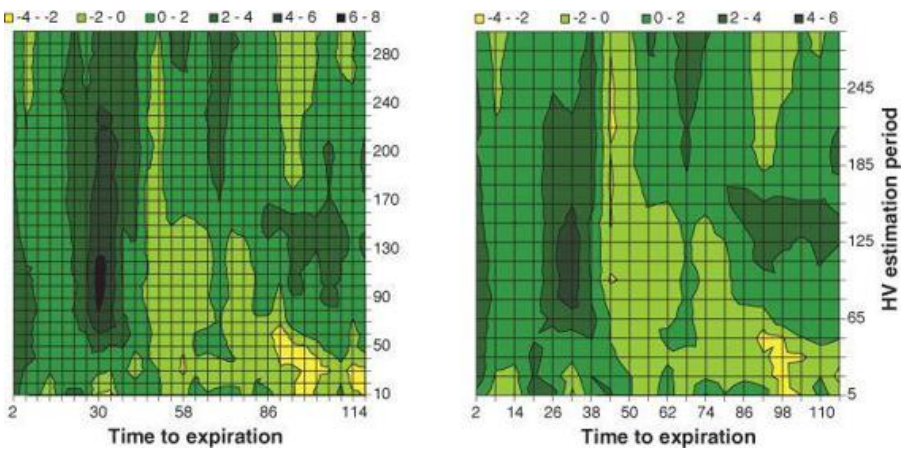

Figure 2.2.4. Optimization spaces of the basic delta-neutral strategy constructed for different optimization steps. Objective function: profit.

Let us keep on increasing the step to 6 days for the "number of days to expiration" parameter and 15 days for the "HV estimation period" parameter. The number of relief details decreases even more (in the right chart of Figure 2.2.4). Furthermore, the original optimal area, which was along the 30th vertical line and contained the global maximum node, has disappeared. The new global maximum has drifted to the node with coordinates 32 and 125. The value of objective function of the new global maximum has declined. This

brings us to the conclusion that *as the optimization step increases, the effectiveness of optimization declines*.

Nevertheless, the increase of the optimization step has its advantages. Despite the shift in global maximum coordinates and worsening of the optimal solution, the new optimal area is still located approximately in the same region of the optimization space where it was originally. At the same time, the surface becomes smoother. The advantage of smoothing is that most insignificant local extremes disappear from the optimization space. Consequently, the probability of the optimization (using more economical methods to search for an optimal solution than the exhaustive search method) stopping at the local maximum decreases.

The following conclusion can be drawn from this study: *By increasing the optimization step, the system developer, on the one hand, decreases the chance that the objective function maximum will get into the optimization space under scrutiny, but, on the other hand, he also reduces the number of computations required and increases search effectiveness by eliminating insignificant local extremes*.

2.3. Objective Functions and Their Application

The objective function is used to evaluate and compare the utility of different combinations of parameter values. That is why selection of appropriate objective function is one of the key elements determining the effectiveness of optimization. Each function creates its own optimization space with specific and sometimes even unique features. Although optimization spaces pertaining to different utility functions may be quite similar in their shape, they may also differ greatly. In this section we analyze objective functions that generate both similar and different spaces in terms of their shape.

In most cases, it is highly unadvisable to limit optimization system to only one utility function. Usually three to four functions must be used simultaneously. Sometimes their number may be even higher (up to ten or more). Applying many objective functions is especially relevant to the optimization of option trading strategies, since in this case not only standard risk and return characteristics have to be evaluated, but also different features that are specific to options. In Chapter 1 we already mentioned several objective functions (when we discussed various characteristics of option portfolios). Using more than one objective function necessitates developing special methods of multicriteria analysis. A substantial part of this section is devoted to this very problem.

2.3.1. Optimization Spaces of Different Objective Functions

The key objective function for optimization of most automated trading systems is profit. This characteristic may be expressed in different forms—in relative (percentage) or absolute figures, annualized or tied to the system's operational cycle. All previous examples in this chapter were based on this objective function. The top-left chart of Figure 2.3.1 reproduces the optimization space constructed for this basic objective function. Now we turn our attention to the following question: In what way does the selection of specific objective functions affect the shape of the optimization space? (The same two parameters that we examined in the previous sections will be used to construct the space.)

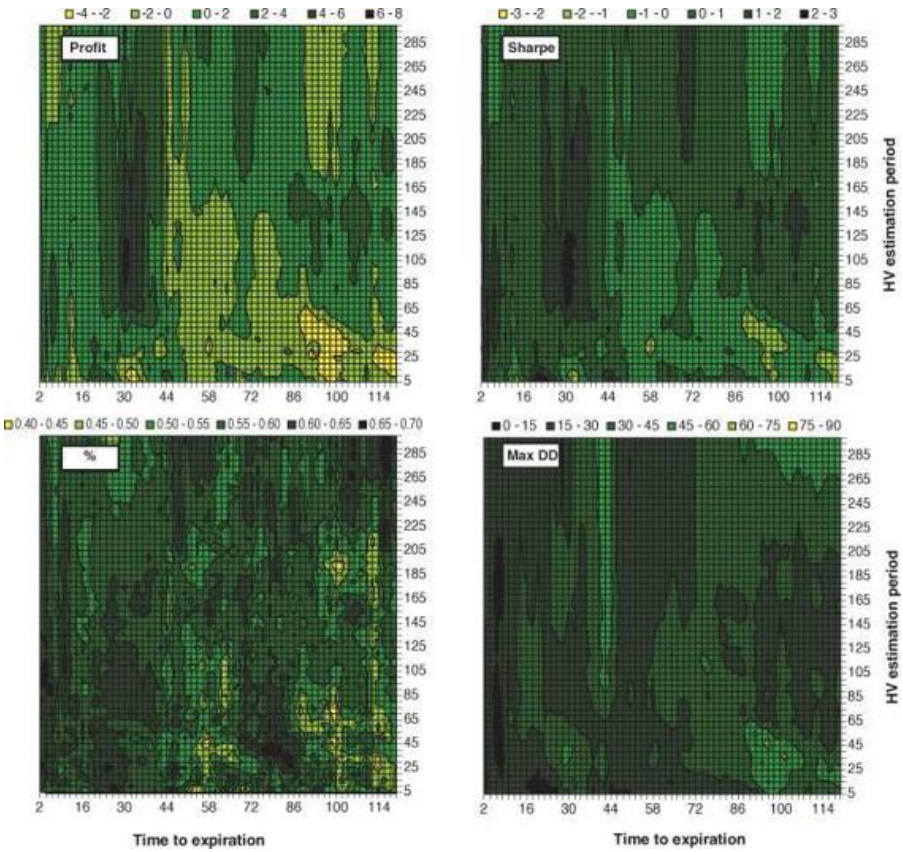

Figure 2.3.1. Optimization spaces of the basic delta-neutral strategy constructed for four different objective functions.

When the Sharpe coefficient is used as the objective function (in the top-right chart of Figure 2.3.1), the optimization space has almost the same appearance as it had for the profit-based function (except for some minor, insignificant discrepancies). The large optimal area is situated in approximately the same place and has a very similar shape. The global maximum has almost the same coordinates as for the

profit-based objective function. The only difference is that the vertical axis coordinate is 100 days for the Sharpe-based function instead of 105 for the profit-based function (this insignificant difference is hardly of any importance). We should also note that this optimization surface is polymodal because it has several other optimal areas with local maximums. However, since these areas are very small, they are inferior to the global maximum area in terms of robustness.

The next objective function, maximum drawdown, represents a risk indicator that is commonly used for evaluation of trading strategies. Optimization space corresponding to this function is shown in the bottom-right chart of Figure 2.3.1. In contrast to other objective functions, optimization based on maximum drawdown requires minimizing the function rather than maximizing it. Thus surface areas with low altitudes are optimal. Contrary to the optimization space constructed using the Sharpe-based objective function, the surface corresponding to maximum drawdown is completely different from the space of profit-based function. Optimal areas are stretched in two directions: (1) in the first half of the values range of the “number of days to options expiration” parameter given that the “HV estimation horizon” has low values (along the horizontal axis); (2) in a wide range of “HV estimation horizon” parameter values given that the “number of days to options expiration” parameter has low values (along the vertical axis). The optimal areas of the Sharpe-based objective function were noted in the same locations of the optimization space (although they were of a much smaller size). This can be explained by the fact that the Sharpe coefficient “contains” risk-related information (standard deviation). Since maximum drawdown is the extreme outlier of the returns time series, it directly influences the standard deviation and indirectly influences the Sharpe coefficient.

The percentage of profitable trades is another indicator applied frequently for measuring the performance of automated trading systems. When this indicator was used as the objective function (in the bottom-left chart of Figure 2.3.1), we obtained the optimization surface that was fundamentally different by its form from the surfaces pertaining to other functions. Firstly, instead of a single optimal area (found for the profit-based objective function) or several such areas (found for the Sharpe-based and maximum drawdown-based functions), the surface of this function is highly polymodal with many optimal areas. Secondly, all optimal areas of this objective function represent small islands, while the majority of areas pertaining to other functions have relatively large surfaces. Thirdly, and this is probably the most valuable feature, optimal areas of other functions do not

coincide with optimal areas of the objective function constructed on the basis of the percentage of profitable trades.

After studying optimization spaces of four objective functions, we can make several important conclusions. *Each function carries a certain amount of important information, part of which is duplicated with information contained in other functions and the rest of which is unique.* The extent to which the information contained in different functions overlaps may be different for different functions. For example, the Sharpe-based function almost does not add new information to the information contained in the profit-based function. Hence, it would hardly be reasonable to use both functions simultaneously (the volume of additional calculations does not justify the little additional information that might be obtained). At the same time, other objective functions do contain a substantial volume of new unique information, which may not be obtained through using the profit-based function. The inclusion of such functions into a multicriteria optimization system might be justified.

In this section we performed a visual (nonformalized) analysis of different objective functions and detected different extents of duplication of the information carried by these functions. In order to express these observations quantitatively—which will enable us to accomplish an objective and unbiased selection of the objective functions—we need to analyze their correlations. This is what the next section is about.

2.3.2. Interrelationships of Objective Functions

In order to express numerically the extent of duplication of information contained in different objective functions, we need to conduct pairwise comparison of their interrelationships. The lower the correlation between two functions, the less the information overlaps and the better it is for multicriteria optimization. The correlation coefficient shows the extent of interrelationship between two functions, and the determination coefficient (the square of correlation coefficient) expresses the extent to which the variability of one objective function is explained by the variability of the second function. Therefore, the portion of additional nonduplicated information that is introduced into the optimization system as the result of including an additional objective function can be expressed by the indicator calculated as one minus the determination coefficient.

Correlation analysis demonstrates that all objective functions are interrelated to a greater or lesser extent (see Figure 2.3.2). As one could expect, the highest correlation was detected between profit and Sharpe coefficient (visual similarities of optimization spaces pertaining

to these two functions have been noted in the previous section). In this case the correlation coefficient is very high ($r = 0.95$). Hence, the portion of nonduplicated information is just about 10% ($1 - 0.95^2 = 0.10$). Therefore, there is no reason to use profit and Sharpe coefficient simultaneously within one optimization system.

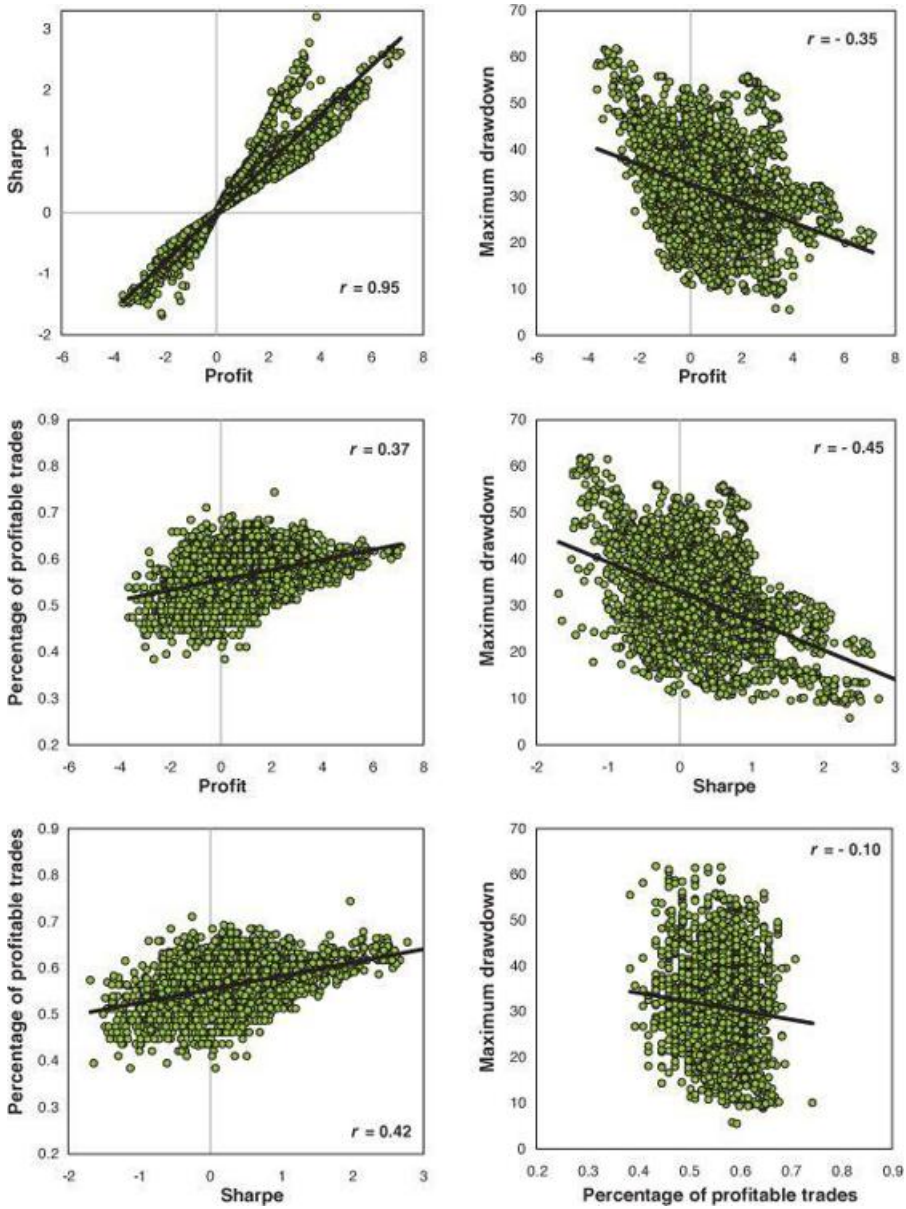

Figure 2.3.2. Correlations of different objective functions.

The extent of interrelationship between profit and the percentage of profitable trades, as well as between profit and maximum drawdown, is much lower than between profit and Sharpe coefficient (an inverse

relationship in the case of maximum drawdown should be treated as a direct one since low drawdown values are preferable). In the first case, the correlation coefficient is 0.37 (in the middle-left chart of Figure 2.3.2), and in the second case it is –0.35 (in the top-right chart of Figure 2.3.2). This means that the portions of nonduplicating information for these pairs of objective functions are 86% and 88%, respectively. These values are high enough to induce thorough consideration concerning the practicability of their inclusion in the multicriteria optimization system. However, before deciding this positively, we need to determine whether it is reasonable to use both additional functions (percentage of profitable trades and maximum drawdown) or whether adding just one of them to the profit-based function would be sufficient.

In order to make this decision, we must analyze the relationship between these two utility functions. As the bottom-right chart of Figure 2.3.2 and low correlation coefficient (–0.10) imply, the percentages of profitable trades and maximum drawdown are almost independent. Information contained in these two functions almost does not overlap (the share of nonduplicating information is 99%). Therefore, adding both objective functions into the multicriteria analysis system is justified.

We can summarize this analysis by stating that it would be reasonable to use three objective functions (profit, percentage of profitable trades, and maximum drawdown) out of four examined candidates. The elimination of the Sharpe coefficient from multicriteria analysis is justified not only by the fact that this function duplicates almost completely the profit-based function, but also by the fact that the Sharpe coefficient correlates with the percentage of profitable trades and maximum drawdown much more strongly than the profit function does (in the middle-right and bottom-left charts of Figure 2.3.2).

Up to this point, the procedure of selecting the objective functions looks quite simple and straightforward. However, we must admit that it was deliberately simplified in order to facilitate description of general principles. Now we remove this simplification to demonstrate the complexity of the objective functions selection process.

The point is that interrelationships shown in Figure 2.3.2 are based on the whole set of data belonging to optimization spaces of corresponding objective functions. This means that these interrelationships were created (and correlations calculated) on the basis of full ranges of two parameter values (2 to 120 days for the number of days to options expiration and 5 to 300 days for the HV estimation period). For example, to estimate the correlation between

profit and Sharpe-based functions, each node in the top-left chart of Figure 2.3.1 was associated with a corresponding node in the top-right chart. The relationship obtained as the result of such association consists of 3,600 points (in the top-left chart of Figure 2.3.2).

However, it would be reasonable to suppose that the extent and even the direction of interrelationships between different objective functions may change depending on specific values possessed by the parameters and, hence, on their value ranges. In order to check this idea, we should calculate correlations separately for each parameter (that is, we need to check whether correlations change depending on any specific values of the parameter).

Let us start with the "HV estimation period" parameter. Correlations between some pairs of the objective functions depend on values of this parameter, and correlations of other pairs of the functions do not (see Figure 2.3.3). For example, the correlation of functions for which the highest extent of interrelationship was observed (profit and Sharpe coefficient) does not change through the whole range of parameter values. All other pairs of functions demonstrate certain patterns.

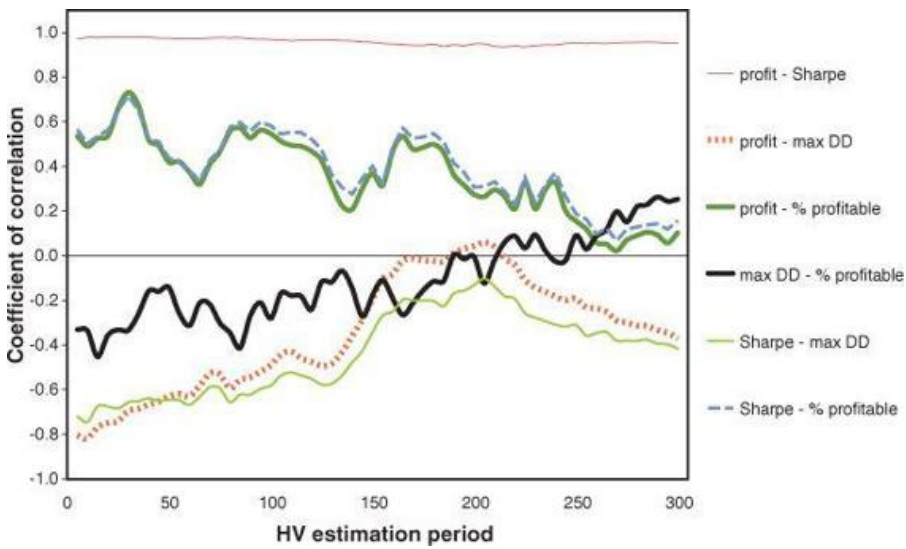

Figure 2.3.3. The relationship between the correlation coefficient of six pairs of objective functions and values of the "HV estimation period" parameter.

The interrelationship between the objective functions expressing profit and percentage of profitable trades is rather high under low values of the "HV estimation period." As parameter values increase, correlation of these two functions decreases, and when the parameter reaches the upper bound of its acceptable values range, it drops to zero. The same trend is observed for interrelationship within another

pair of objective functions—Sharpe coefficient and the percentage of profitable trades. This similarity in trends is not surprising given that profit and Sharpe coefficient are almost perfectly correlated.

Two other pairs of objective functions (profit and maximum drawdown; Sharpe coefficient and maximum drawdown) also demonstrate almost identical trends (the reason for similarity of the trends is the same as in the previous case). Correlations for these two pairs are quite high under low parameter values (remember that in case of maximum drawdown, negative correlation coefficient has the same meaning as positive correlation in other cases). As the parameter increases to average values, correlations approach zero and then become negative again. When the same data were examined aggregately (in the top-right and middle-right charts of Figure 2.3.2), we could discover only the inverse relationship for these pairs of objective functions. The detailed analysis presented in Figure 2.3.3 reveals that under average values of the "HV estimation period" parameter, there are no correlations at all within these two pairs of objective functions (and hence information contained in them is not duplicated).

The interrelationship between the last pair of objective functions (maximum drawdown and percentage of profitable trades) demonstrates the uptrend. Under low parameter values the correlation coefficient is negative, and when the parameter reaches its highest values, correlation becomes positive (the correlation coefficient is zero when the HV estimation period is about 200 days). Note that when the same data were considered aggregately (in the bottom-right chart of Figure 2.3.2), we could not detect any relationship between these objective functions. As a result, we could come to a wrong conclusion that information carried by these functions is completely nonduplicating. Meanwhile, the detailed analysis (see Figure 2.3.3) specifies that the information is nonduplicating only in the second third of the parameter values range.

Now we go on to consider the influence of the second parameter (number of days to options expiration) on the interrelationship between objective functions (see Figure 2.3.4). This parameter influences correlations of objective functions much stronger than the "HV estimation period" parameter does (compare Figure 2.3.3 and Figure 2.3.4). Even small changes in the parameter lead to significant changes in correlations. Correlation coefficients of almost all pairs of objective functions fluctuate in a very wide range (from –0.9 to 0.9). However, in contrast to the previous case (when we examined the effect of the "HV estimation period" parameter), the influence of the number of days to options expiration is rather chaotic. The dynamics

of correlation coefficients have no distinguishable trends.

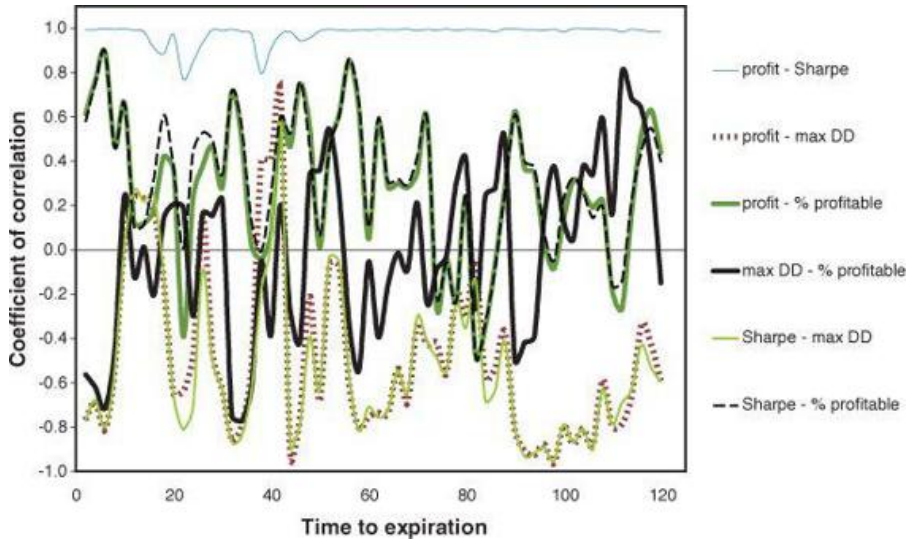

Figure 2.3.4. The relationship between the correlation coefficient of six pairs of objective functions and values of the "number of days to options expiration" parameter.

The pair of objective functions expressing profit and Sharpe coefficient presents the only exception. In this case the correlation coefficient does not depend on the number of days left to options expiration and is persistently at its maximum almost at the whole range of parameter values (see Figure 2.3.4). The same relationship was observed for this pair of objective functions when the influence of the "HV estimation period" parameter was analyzed.

In general, we can conclude that when deciding to include or not include a given objective function in the multicriteria optimization system, the interrelationships between this function and other candidate functions must be considered. Functions with the lowest possible correlations should be preferred. This will introduce new nonduplicated information into the whole system. While establishing the extent of interrelationships that is permissible for accepting an objective function (the threshold for the correlation coefficient below which the objective function is accepted), we have to make sure that the correlation used in the decision-making process is independent of parameter values. If such dependence does exist (as it was shown previously), we should use the correlation coefficient that has been obtained on the basis of data corresponding most closely to the logic of the trading strategy under development.

2.4. Multicriteria Optimization

In the preceding section we discussed selection of objective functions for multicriteria optimization. This section is devoted to the search of optimal solutions using multicriteria analysis methods. As applied to parametrical optimization, the purpose of multicriteria analysis consists in simultaneous utilization of many objective functions (each of which represents a separate criterion) to range nodes of a given optimization space (each node represents a unique combination of parameter values) by the extent of their utility.

The main problem of multicriteria optimization is that complete ordering of all alternatives may turn out to be impossible due to their nontransitivity. Let us illustrate this with a simple example. Assume that an alternative is regarded as the best one if it outmatches all other alternatives by most criteria. Suppose that the comparison of three nodes (*A, B,* and *C*) by values of three objective functions (criteria) gives the following result: $A = (1; 2; 3)$, $B = (2; 3; 1)$, $C = (3; 1; 2)$ (criteria values are given in parentheses). Obviously, node *B* is preferred to node *A* according to the values of the first and second criteria. *C* is better than *B* by the first and third criteria. If transitivity is maintained, it should follow that node *C* is preferred to *A*. However, that is not the case, since *A* surpasses *C* by two criteria, the second and the third ones.

Unfortunately, the nontransitivity problem has no universal solution. Nevertheless, there are two main approaches to obtaining an optimal solution (or several solutions) despite the failure to comply with transitivity. The first approach is based on reducing all objective functions to the single criterion called "convolution." The second approach consists of application of the Pareto method.

2.4.1. Convolution

Giving up simultaneous application of several criteria and substituting them with a new and only criterion (which is a function with initial criteria serving as arguments) constitutes the approach called "convolution." The advantage of convolution is the simplicity of realization and the possibility to adjust the extent of influence of each criterion on optimization results. This can be done by multiplying criteria values by weight coefficients—the higher the weight of a particular criterion, the greater its impact on the final result of the multicriteria optimization. The main drawback of convolution is an unavoidable loss of information that occurs when many criteria are transformed into a single one.

Two convolution types are used commonly: additive (a sum or an arithmetic mean of parameter values) and multiplicative (a product or geometrical mean of parameter values). Application of multiplicative

convolution is possible only if both criteria are positive (since the product of two negative values is positive), or if only one of the criteria may take on a negative value. One must also take into account that when one of the criteria is zero, the multiplicative convolution also equals zero regardless of the value possessed by the second criterion (it is not the case for additive convolution). Criteria with lower values have more influence in the multiplicative than in the additive convolution. Additive convolution is more appropriate for criteria having values with similar scales.

Apart from additive and multiplicative convolutions there is a *selective convolution* when the lowest (the most conservative approach) or the highest (the most aggressive approach) value is taken from the whole set of objective functions as a convolution value for each node. In our previous book (Izraylevich and Tsudikman, 2010) we introduced *minimax convolution,* for which the product of the highest and the lowest criteria is used as the convolution value.

When calculating the convolution value, we must remember that criteria may be measured in different units and have different scales. The following transformation can be used to reduce them to the uniform scale with the same value ranges:

$$x_{transformed} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \qquad (2.4.1)$$

where x_i is the criterion value for the i-th node, and x_{min} and x_{max} are minimum and maximum criterion values, respectively. Application of this formula reduces the criterion value to the range from 0 to 1.

Let us consider an example of applying the convolution concept to the basic delta-neutral strategy. Suppose that three of the four objective functions presented in Figure 2.3.1 were selected to serve as criteria—profit, maximum drawdown, and the percentage of profitable trades (Sharpe coefficient is rejected due to its high correlation with profit). In the first place, values of all three objective functions must be reduced to the range 0 to 1 (by applying formula 2.4.1). Having created three types of convolutions (additive, multiplicative, and minimax), we observed that in this particular case they all produced rather similar results.

Figure 2.4.1 shows the optimization space of minimax convolution. This surface is polymodal and has four optimal areas. Three of them have a relatively vast area and one is very small (since the surface area is one of the main factors in determining the selection of optimal solution, the fourth area may be excluded from our consideration). In this case, multicriteria analysis based on the convolution of objective functions did not allow establishing a single optimal solution since

each of the three areas has its own local maximum. Hence, the application of convolution did not solve the optimization problem completely. We still have to select one of these three areas. Since all of them have approximately the same altitudes (convolution values), the selection must be based on a different principle. In the next sections we discuss the selection of an optimal area on the basis of its relief characteristics and quantitative estimates of robustness.

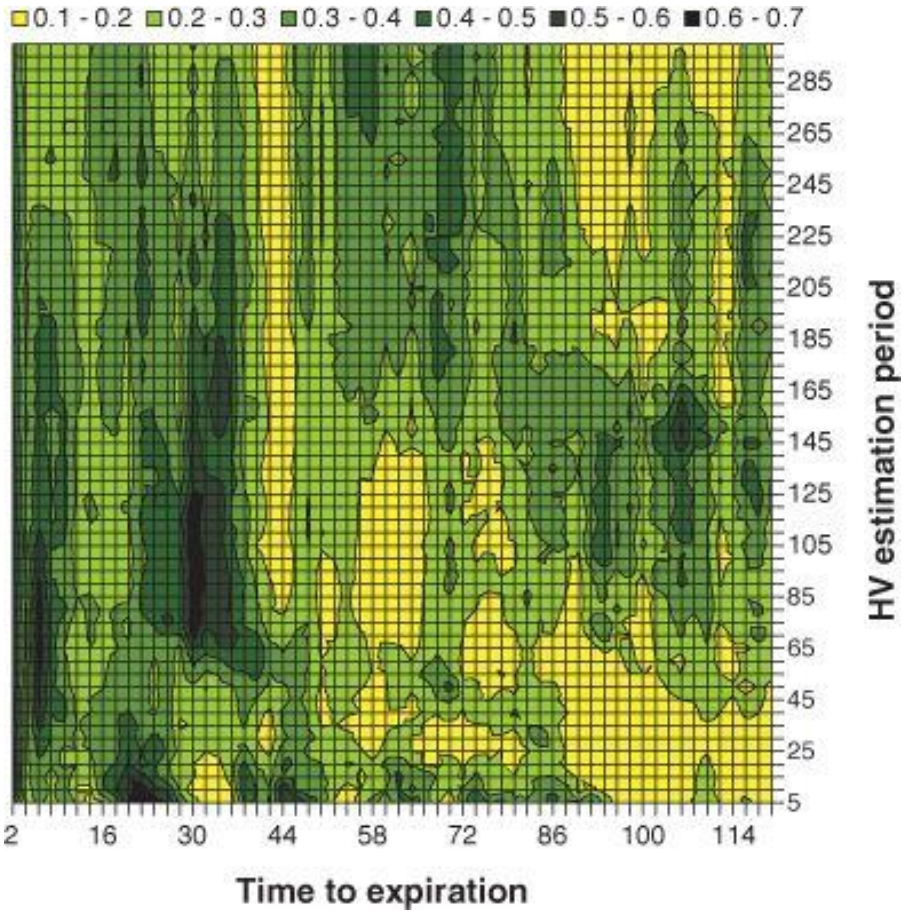

Figure 2.4.1. Multicriteria optimization of the basic delta-neutral strategy. Optimization space is created by constructing a minimax convolution of three objective functions (profit, maximum drawdown, and the percentage of profitable trades).

2.4.2. Optimization Using the Pareto Method

Application of the Pareto method allows solving the selection problem when different criteria values contradict each other. The situation, when some nodes of the optimization space are better than others according to the values of one objective function (for example, profit), but are worse according to another function (for example,

maximum drawdown), is rather frequent. The main drawback of the Pareto method is that we can get several equally good solutions instead of a single one. In such circumstances the selection of a unique solution will require further analysis and application of additional techniques.

Formalization of the multicriteria optimization problem is realized in the following way. Assume that for each node (alternative) a of the optimization space there is n-dimensional vector of objective functions (criteria) values $x(a) = (x_1(a),...,x_n(a))$. Using values of n criteria, we need to find alternatives with maximum values of vector coordinates (that is, with maximum values of corresponding objective functions). Suppose that the higher the criterion value, the better the alternative. When alternative a is compared to alternative b, alternative a is deemed to dominate over alternative b if the following set of inequalities holds: $x_i(a) \geq x_i(b)$, for all $i = 1,...,n$, and there is at least one criterion j for which strict inequality $x_j(a) > x_j(b)$ holds. In other words, node a is preferred to node b if a is not worse than b according to the values of all objective functions and is better than b by at least one of them.

Obviously, the presence of domination (in the sense defined previously) determines unambiguously which of the two alternatives is better. However, if domination is uncertain, the question of which alternative is a better one remains open. In this case we have to admit that none of the alternatives is unambiguously superior to the other one.

Based on the previously stated reasoning, the multicriteria optimization problem can be defined in the following way: *Among the set of all possible alternatives, identify the subset that includes only nondominated alternatives, that is, those for which there are no alternatives dominating over them.* This subset is the Pareto set. Each element belonging to it can be considered the best in the sense defined previously. The number of alternatives composing the Pareto set may be different. It may be one alternative dominating over all the rest, several best alternatives, or even the whole initial set.

In our example of the basic delta-neutral strategy, the optimization space $A = \{a_1,...,a_m\}$ consists of m alternative nodes ($m = 3{,}600$ in the example), evaluated with n functions-criteria ($n = 3$) with values $x(a) = (x_1(a_1),...,x_n(a_m))$. To establish the Pareto set, we need to arrange a pairwise comparison of all the alternatives, during which we get rid of dominated ones and include nondominated variants into the Pareto set. Each alternative a_k is compared with all the rest. If there is the alternative a_l that is dominated by a_k, a_l is discarded. If a_k is dominated by some element a_m from the remaining elements, a_k is

discarded. If none of the elements dominates over a_k, the latter is included in the Pareto set. Afterward, we pass on to comparing the next variant with all the rest. The maximum number of comparisons required is about $0.5 \times m \times (m\text{-}1)$, which is quite reasonable in most cases. However, faster algorithms may be needed to build the Pareto set when more criteria and alternatives are involved in the optimization procedure.

As mentioned earlier, the drawback of the Pareto method is the impossibility to influence the number of nodes included in the Pareto set. The number of alternatives belonging to the optimal set may change from case to case and does not depend on our wishes and preferences. The unique optimal solution may be obtained only when the optimization space has the node for which all criteria values are higher than their values for any other nodes. However, in most cases, several optimal solutions, rather than a single one, are present.

Let us consider the application of the Pareto method for optimizing the basic delta-neutral strategy. The same three objective functions that were used in multicriteria optimization with the convolution method (profit, maximum drawdown, and percentage of profitable trades) will be used again as the criteria. In contrast to convolution, the Pareto method does not allow constructing an optimization space similar to the surface shown in Figure 2.4.1. Instead, we get the list of dominating nodes that compose the optimal set. Consequently, the optimization surface turns into the coordinate space designating the position of separate optimal nodes (see Figure 2.4.2).

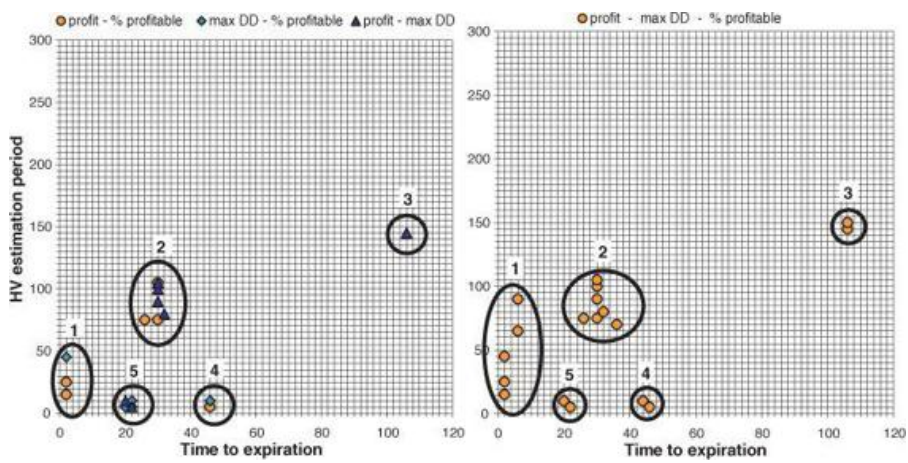

Figure 2.4.2. Multicriteria optimization of the basic delta-neutral strategy by the Pareto method. The left chart shows optimal sets obtained by applying two objective functions (three optimal sets corresponding to three different function pairs are shown). The right chart shows the optimal set obtained by

simultaneous application of three objective functions.

First we consider the Pareto set obtained by applying two criteria. Three objective functions can give rise to three criteria pairs, each of which produces its own optimal set. Nodes composing these optimal sets are grouped in several areas in the coordinate space. On the left chart of Figure 2.4.2, these areas are denoted with sequence numbers. Interestingly, none of these areas contains nodes belonging to all three optimal sets. The third area contains a single node belonging to the set obtained by applying profit-based and maximum drawdown objective functions. Nodes selected by this pair of functions are also present in the second and fifth areas. Nodes corresponding to the pair of objective functions that are based on profit and percentage of profitable trades are found in the first, second, and fourth areas. Finally, nodes contained in the optimal set, selected using objective functions based on maximum drawdown and percentage of profitable trades, are located in the first, fourth, and fifth areas. Such distribution of optimal sets in coordinate space indicates that each of the three objective functions makes its own contribution to solving the optimization problem. Therefore, in this case it would be reasonable to include all three objective functions in the Pareto system of multicriteria optimization.

The set of optimal solutions obtained as a result of applying three criteria simultaneously is shown in the right chart of Figure 2.4.2. Nodes belonging to the optimal Pareto set are located approximately in the same five areas as in the previous example when criteria were applied by pairs. The highest number of nodes (seven in total) is situated in the second area; the first area contains five nodes; and the third, fourth, and fifth areas contain only two nodes each.

In the previous example, when criteria were used by pairs, optimal sets contained five to seven nodes (depending on the specific criteria pair). Simultaneous application of three criteria enlarged the set to 18 elements. This is the general property of the Pareto method—increasing the number of criteria leads to increasing the number of optimal solutions, other things being equal. Since effective resolution of the parametric optimization problem requires selecting a single optimal node, including each additional utility function into the optimization system complicates this task significantly. Thus, when making a decision concerning the inclusion of a particular objective function into a specific optimization framework, we should not only consider the scope of the new information added to the system by each additional function, but also account for increasing complexity of the selection of a single optimal solution.

Distribution of optimal areas obtained by applying the convolution method is quite similar to the distribution observed when optimal solutions were determined using the Pareto method (compare Figure 2.4.1 and Figure 2.4.2). This means that both methods lead, in this particular case, to approximately the same results. However, it should be noted that under other circumstances, application of these methods might bring about completely different outcomes.

The most important conclusion that follows from multicriteria optimizations considered in this section is that both methods define more than one optimal area. None of them can be designated as the unique optimal solution. Consequently, we must admit that the optimization problem has not been solved completely. We should continue the search for the unique optimal solution. This issue is discussed in the next section.

2.5. Selection of the Optimal Solution on the Basis of Robustness

As shown in the preceding section, multicriteria optimization has one essential drawback. In most cases simultaneous application of several objective functions leads to selection of several optimal solutions. While none of them is better than others, we still need to choose one unique alternative. Such a choice can be made by considering the shapes of the optimal areas that surround optimal nodes.

When any particular node is evaluated in the process of multicriteria analysis, only the values of the objective functions that correspond to this particular node are taken into account. Objective functions of adjacent nodes are totally ignored. Meanwhile, the relief of the optimal area is an important indicator of optimization reliability. All other things being equal, *the node is preferred if it is situated near the center of a relatively smooth high area* (the height is determined by the values of objective functions). It is also desirable that this area has wide, gentle slopes. This means that nodes surrounding the node of optimal solution should be close to it by objective function values.

According to the definition given earlier, an optimal solution situated within such an area is robust. If the optimal solution is situated in the area with variable altitudes, sharp peaks, and deep lows, it is less robust and, consequently, less reliable.

Although robustness has no strict mathematical definition with regard to optimization procedure, we can state that an optimal solution located on the smooth surface is more robust than a solution

on the broken surface. If the solution is robust, small changes in values of optimized parameters do not lead to significant changes in the values of objective functions.

In order to base the selection of the optimal solution not only on the altitude mark, but on robustness as well, it is necessary to evaluate quantitatively the relief of the optimal area and the extent of its roughness. When optimization space is multidimensional, this problem is very complex and requires application of topology methods. However, for two-dimensional spaces we can propose several relatively simple methods.

2.5.1. Averaging the Adjacent Cells

This method of robustness evaluation is similar to the concept of moving averages. When the moving average is calculated, values of the objective function (usually price or volume) are averaged within a certain time frame that is moving in time. This procedure is used to describe time dynamics and to determine trends in price or other indicators. For analyzing the relief of optimization space and evaluating the robustness of the optimal solution, averaging of the objective function is performed while moving through the optimization space. In each node of the space, the value of the objective function is substituted with the average value of the objective functions of a small group of adjacent nodes. Thus, the original optimization space is transformed into a new space that will subsequently be used to find the optimal solution. The search in this transformed space is performed as usual, by the altitude marks. The new altitude mark of each node contains information not only about the objective functions of the node itself, but also about the objective function values of a small area surrounding this node. Therefore, applying this method enables us to take into account the robustness of the optimal solution.

The only averaging parameter is the size of the area whose nodes are averaged. These may be only adjacent nodes (one line of nodes surrounding a given node). If the optimization space is two-dimensional, each node is adjacent to eight other nodes (excluding nodes situated at the borders of acceptable parameter values). Thus, when one line of nodes is averaged, the average is calculated on the basis of nine values—eight values of adjacent nodes plus the value of the central one. In averaging two lines, the calculation is performed with 25 values; for three lines, 49 values; and so on. In general, the number of averaged nodes n is computed as

$$n = (2m + 1)2,$$

where m is the number of lines of nodes surrounding the central

node.

Application of this procedure can be demonstrated using the optimization space obtained earlier as the convolution of three objective functions. The original optimization space (see Figure 2.4.1) contains three optimal areas, each of which may be selected to perform the final search for a single optimal solution. Figure 2.5.1 shows two transformations of the original optimization space: The left surface was created using $m = 1$ (averaging one line of adjacent nodes), while for the right surface $m = 2$ (two lines are averaged). After the transformation, based on the averaging of only the closest nodes (one line), just one optimal area remained out of the three that were presented originally. It is located in the range of 28 to 34 days by the "number of days to expiration" parameter and from 75 to 125 days by the "HV estimation period" parameter. The reason for the disappearance of the two other areas is that their extreme values turned out to be less robust than the extremum of the area that had persisted. The transformation obtained by averaging more nodes (in the right chart of Figure 2.5.1) generates similar results—the disappearance from the optimization space of two less robust areas and preservation of just one optimal area. Thus, both transformations point to the same preferable area. Besides the highest robustness, this area has also the biggest surface (see the original optimization space, in Figure 2.4.1). This additional advantage supports the decision to select the optimal solution within this particular area.

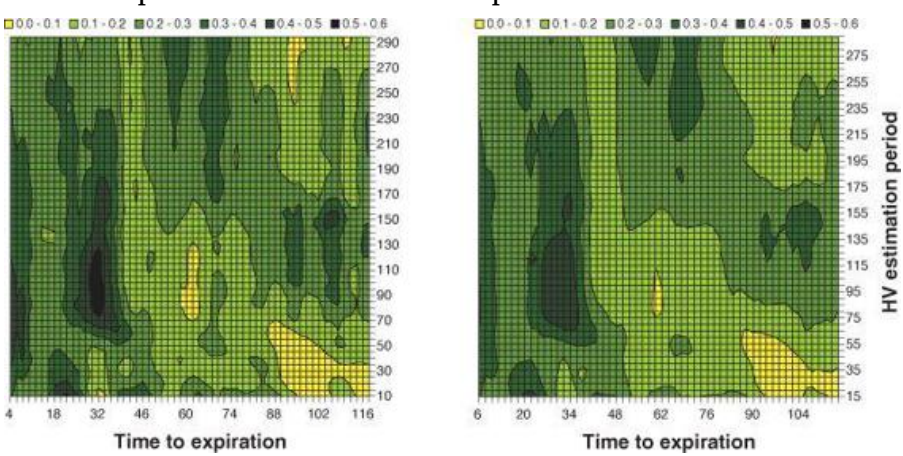

Figure 2.5.1. Transformation of the optimization space of the convolution shown in Figure 2.4.1 by application of the averaging method. The left chart is obtained by averaging one line of adjacent nodes; the right chart, by averaging two lines.

2.5.2. Ratio of Mean to Standard Error

The averaging method described in the preceding section takes into

account the height (objective function value) and smoothness (robustness) of the optimal area, but the influence of the former characteristic is greater than that of the latter one. The method proposed in this section assigns much greater weight to the robustness. According to this method, the objective function value in each node of the original optimization space is substituted with the ratio of the mean of objective function values, calculated for a group of surrounding nodes, to the standard error, calculated for the same group. The notion of the "group of nodes" has the same meaning as in the averaging procedure. The group contains the central node itself and one, two, or more lines of the adjacent nodes. Such transformation of optimization space takes into account both the altitude of the optimal area (numerator) and the smoothness of its relief (denominator).

Figure 2.5.2 demonstrates two transformations of the convolution shown in Figure 2.4.1. As in the preceding section, transformations were constructed using one and two lines of adjacent nodes ($m = 1$ and $m = 2$, respectively). When one line of nodes was used for transforming the original optimization space (in the left chart of Figure 2.5.2), many new optimal areas arose (the original space contains only three such areas). The number of new areas is so high that objective selection of one of them is practically impossible. This problem can be solved by using a larger group of nodes (two lines). This significantly simplifies the relief of the transformed optimization space (in the right chart of Figure 2.5.2). Only three optimal areas (from which we need to choose one) are present now.

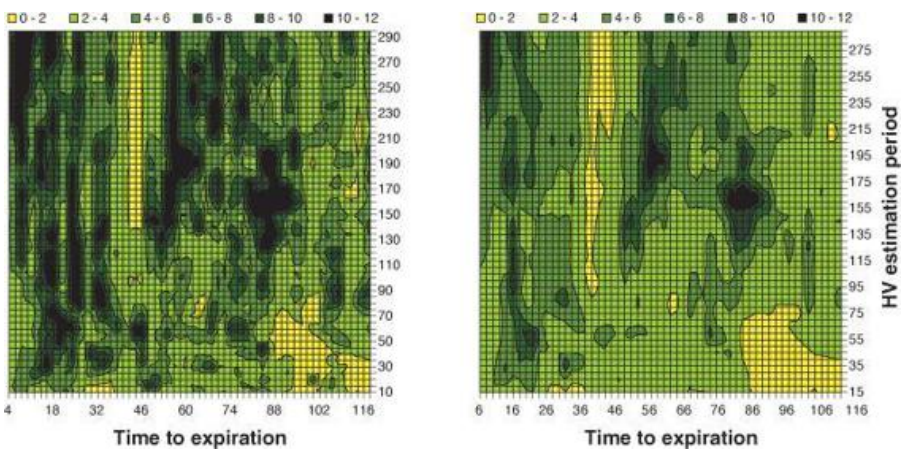

Figure 2.5.2. Transformation of the optimization space of the convolution shown in Figure 2.4.1 performed using the ratio of mean to standard error. The left chart is obtained by using one line of adjacent nodes; the right chart, by using two lines.

In contrast to the transformation obtained by simple averaging (see the preceding section), location of none of the three optimal areas coincides with locations of optimal areas on the original optimization space. The reason is that optimal areas of the original convolution represent narrow ridges and high peaks. These areas are not robust enough and their relief is quite broken (hence, they disappear as a result of the transformation). As opposed to this, optimal areas of the transformed optimization space represent fairly smooth and wide plateaus of average height, which might be preferable from the robustness point of view. Since the new optimal areas are almost equivalent in respect to both the objective function values and the robustness, we can select one of them by the size of the surface area and its shape. Other things being equal, it is better if the optimal area has a greater surface area and a more rounded shape (narrow areas are less robust at least by one of the parameters). These criteria are satisfied for the area situated in the range of 84 to 92 days by the "number of days to expiration" parameter and 155 to 175 days by the "HV estimation period" parameter.

2.5.3. Surface Geometry

Two methods described in the previous sections are based on transformation of the optimization space. Now we move to describing the alternative approach that does not require transformation. It is based on analyzing and comparing the geometry of optimal areas. Actually, the developer of a trading system can develop a multitude of similar methods employing different mathematical techniques of any complexity level. Here we propose one possibility to solve the problem of selecting the optimal solution while taking into account its robustness.

To explain the main principles of the proposed method, let us consider the hypothetical optimization space (see Figure 2.5.3) containing two optimal areas. One of them (located in the left part of the space) has a smaller surface area, a higher top, and a more stretched form. Another area (located in the right part of the space) has a bigger surface, a lower top, and a more rounded form. We need to decide which of them is preferable for optimal solution searching. The approach we propose is based on the assumption that the optimal area with greater space area is preferable. By "space area" we mean the area of an elevation measured in the three-dimensional space (in the case of two-dimensional optimization), rather than surface area of its two-dimensional projection. This assumption is quite realistic since a greater area may indicate a superior combination of two important characteristics—higher value of the objective function at the extreme point and higher robustness of the potential optimal solution.

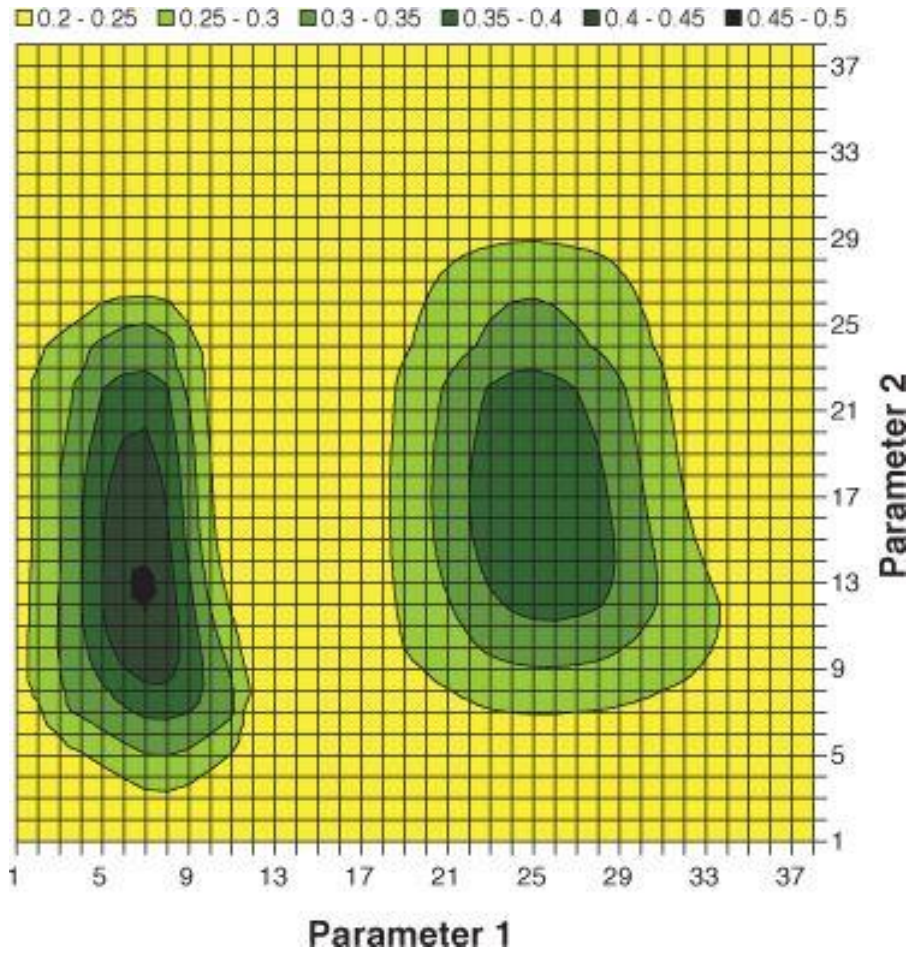

Figure 2.5.3. Hypothetical example of the optimization surface containing two optimal areas with different shapes.

In the following discussion we assume that the optimal area represents a cone. Undoubtedly, this assumption is a simplification; real areas have more complex shapes. Nevertheless, in reality the shape of many optimal areas resembles a cone. Any optimal area may be reduced to a cone by simple mathematical manipulations. This allows calculating the surface area without applying complex differentiation methods. In order to estimate the surface area by reducing a certain part of optimization space to a cone, the following procedures should be implemented.

1. Specify the level of the objective function that determines the border of the optimal area. All nodes situated above this altitude mark are considered to belong to the optimal area and, correspondingly, to the cone surface. This level circumscribes the base of the cone. The level accepted for the

theoretical example depicted in Figure 2.5.3 is 0.25.

2. Calculate the area of the cone base as the number of nodes located within a given optimal area. In the example presented in Figure 2.5.3, this area is equal to the number of nodes within the border that encompass the nodes with altitude marks exceeding the 0.25 level.

3. Knowing the cone base area of k, calculate the radius of the symbolic circle at the cone base as $r = \sqrt{k/\pi}$.

4. Calculate the area of the lateral cone surface as $S = L\pi r$, where L is the cone side.

Using the Pythagorean theorem, we calculate the side of the cone as

$$L = \sqrt{h^2 + r^2} = \sqrt{h^2 + \frac{k}{\pi}},$$

where h is the cone height. The cone height is known—it is the value of the objective function at the extreme point of the optimal area. Knowing the side length, we can calculate the area of the side surface of the cone as

$$S = \pi\sqrt{\frac{k}{\pi}}\sqrt{h^2 + \frac{k}{\pi}}.$$

After simple algebraic transformations we get

$$S = \sqrt{k\pi h^2 + k^2}. \qquad (2.5.1)$$

It is important to note that k and h are expressed in different units. The first one represents the number of nodes, and the second one is the objective function value, which may take various forms (percent, dollars, or any other). That is why S is a dimensionless indicator, which is not an area in its original sense, though it is proportional to the real area of the optimal surface and can be used to compare different optimal areas. To avoid misunderstanding, we will further refer to this characteristic as a "conventional area." It is necessary that k and h have approximately the same magnitude (for example, if $h = 5$ and $k = 70$, we should adjust k by dividing its value by 10). Besides, h must not be less than 1; otherwise, squaring it will not increase but decrease the final result (see formula 2.5.1).

Let us demonstrate the practical application of this method. In the hypothetical example presented in Figure 2.5.3, the left area consists of 185 nodes; the objective function value of the node with the highest altitude mark is 0.47 (that is, $k = 185$, $h = 0.47$). For the right area $k = 266$ and $h = 0.40$. To reduce k and h to the same scale, we divide k by 100 and multiply h by 10 (to satisfy $h > 1$). Using

formula 2.5.1, we get

$$S_{left} = \sqrt{1.85 \times 3.14 \times 4.7^2 + 1.85^2} = 11.48, \qquad S_{right} = \sqrt{2.66 \times 3.14 \times 4^2 + 2.66^2} = 11.86.$$

The conventional area of the left optimal area is smaller than that of the right one. This means that despite the fact that the left area is higher, the right one should be preferred. Therefore, in this case the advantages of the wider and more sloping surface of the right area (that is, the advantage of robustness) outweighed the advantage of the higher objective function value of the left area.

Now we turn our attention to the example based on the real market data. The right chart of Figure 2.5.2 contains three areas with altitude marks higher than 10 (remember that this optimization space represents transformation of the initial space obtained by convolution of three utility functions). We denote these areas as «left», «middle» and «right» ones. All three areas have similar base areas ($k_{left} = 10$, $k_{middle} = 13$, $k_{right} = 14$) and heights ($h_{left} = 14.09$, $h_{middle} = 13.45$, $h_{right} = 11.91$), which turns the selection of one of them into a difficult task. However, by applying the method based on the surface geometry, we can make an objective choice. Since values of k and h are scaled similarly and $h > 1$, no adjustments are required. Substituting values into formula 2.5.1, we get $S_{left} = 79.6$, $S_{middle} = 86.9$, $S_{right} = 80.2$. On the basis of these conventional area estimations, we conclude that the middle area is preferable. This decision is not trivial since the middle area has neither the biggest base area nor the highest altitude mark. This example demonstrates combined application of two methods. First, the optimization space was transformed by calculating the ratio of the mean to the standard error, and then the selection of optimal area was performed on the basis of surface geometry.

2.6. Steadiness of Optimization Space

In the preceding section we compared optimal areas of the optimization space on the basis of their robustness. We defined robustness as the sensitivity of the objective function to small changes in values of the parameters under optimization. The desired property of the optimal solution is high robustness (that is, nonsensitivity to parameter changes). It is important to emphasize that in this context we talk about the robustness relative to *optimized parameters.*

In this section we will discuss another aspect of robustness: the extent of optimization space sensitivity to *fixed (that is, non-optimized) parameters.* During optimization based on algorithmic calculation of the objective function value, a multitude of combinations of optimized parameter values are analyzed. Still, many other parameters are kept

unchanged (we call them “fixed parameters”). Their values may be selected at earlier stages of strategy development (using the scientific approach), or they may be defined by the strategy idea. *The robustness of the optimization space relative to small changes in values of the fixed parameters and to inessential alterations of the strategy structure is an important indicator of optimization reliability.* When referring to the sensitivity of the optimization space to any changes other than changes in optimized parameters (which are measured by robustness), we will use the term “steadiness.”

The following reasoning would explain this idea. During optimization of the basic delta-neutral strategy, we have obtained a certain optimization surface (see Figure 2.2.2). In this specific case only two parameters have been optimized, while values of all others were fixed. In particular, the “criterion threshold” parameter was fixed at 1% (positions opened only for option combinations with expected profit exceeding 1%). Suppose that we decided to increase the value of this parameter to 2%. Assume further that this change led to a transformation of the original optimization surface (and now its appearance is similar to that shown in Figure 2.5.2). If such a small change of the fixed parameter led to such a significant surface change, we would conclude that this surface is unsteady, the optimization is unreliable, and hence it is very risky to trust its results.

2.6.1. Steadiness Relative to Fixed Parameters

In this section we analyze the steadiness of the optimization space of the basic delta-neutral strategy relative to the fixed “criterion threshold” parameter. Figure 2.2.2 shows the surface obtained using the profit-based objective function provided that the criterion threshold is fixed at 1%. We will increase the value of this parameter to 3% and test the consequences of this change on the shape of the optimization surface.

Before the fixed parameter was changed, the global maximum had coordinates of 30 by the “number of days to expiration” parameter (horizontal axis of the chart) and 105 by the “HV estimation period” parameter (vertical axis). When we increased the value of the fixed parameter, the global maximum shifted to the node with coordinates of 16 and 120, correspondingly. The magnitude of this shift cannot be regarded as significant (though it is not negligible either).

The original optimization space had a single optimal area along the 30th vertical in the range of 80 to 125 days by the “HV estimation period” parameter (see Figure 2.2.2). The left chart of Figure 2.6.1 demonstrates the new optimization space resulting from the fixed parameter change. The former optimal area remained almost at the

same place (shifted down a bit) and slightly increased in size. At the same time, four new optimal areas emerged to the left of the original one; two of them are very small and another two are comparable to the previous one in size. It is important to note that although the number of optimal areas increased (five instead of one), all of them are situated around the bottom-left part of the optimization space (12 to 36 days by the "number of days to expiration" parameter and 40 to 180 days by the "HV estimation period" parameter).

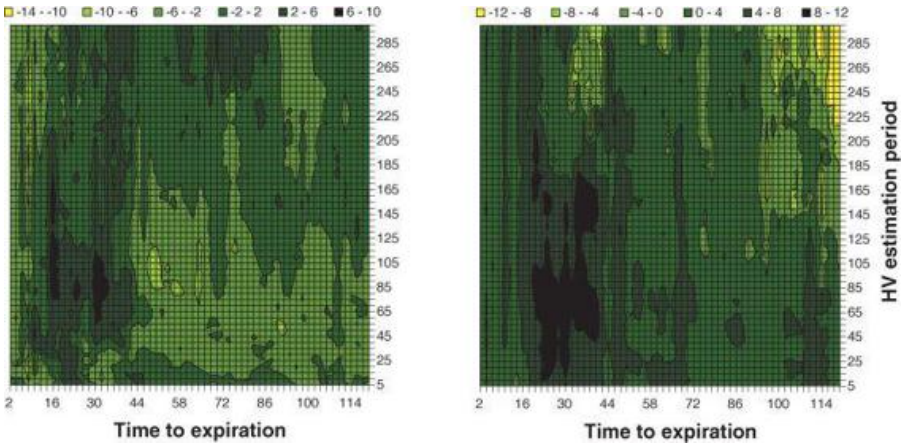

Figure 2.6.1. The change of the optimization space of the basic delta-neutral strategy (the original space is shown in Figure 2.2.2) induced by the change in value of the "criterion threshold" parameter (left chart) and by prohibition of long positions (right chart).

We can conclude from these data that the change in the value of this fixed parameter did not change the shape of the optimization space fundamentally. It testifies to the relative steadiness of the optimization space. Although the changes that have taken place may seem to be quite significant, it should be taken into account that the change in the value of the fixed parameter (from 1% to 3%) was also substantial. We used this big change on purpose to make a clear demonstration of the change in the optimization space. When testing the steadiness of the automated strategy intended for real trading, it would be sufficient to apply much smaller changes in values of fixed parameters.

2.6.2. Steadiness Relative to Structural Changes

The concept of structural steadiness is very broad. The strategy structure includes various aspects from the basic idea to relatively insignificant technical elements. The change of some structural elements may change the shape of optimization space radically. The

developer of the trading strategy should strive to obtain the optimization space, which is relatively steady to *small changes in the strategy structure*. The point is that any significant change turns the strategy under optimization into an entirely different strategy that may require an application of a completely different approach to optimization.

Small structural changes may include slight altering of a capital allocation method or application of an alternative risk management instrument. With regard to the delta-neutral strategy, the change in the method used to calculate index delta may be considered as a structural change. On the one hand, changes in these and similar structural elements do not change the nature of the strategy. On the other hand, it is highly desirable that the optimization space would demonstrate steadiness to changes of such kind. Undeniably, if the scheme of capital allocation among portfolio elements changes, but the shape of corresponding optimization space does not change significantly, such optimization can be regarded as steady and reliable. One should also note that the same optimization space can be steady to some structural changes and have a low steadiness to others.

Let us consider an example of another structural change—limitations of long positions. In its extreme form (complete prohibition to open long positions) such a change is rather considerable and may even completely change the nature of a trading strategy. However, this restriction must not be absolute. For example, a "soft" limitation may be applied that restricts the share of long positions in the portfolio to the predetermined threshold level. To make the change of the optimization space more spectacular and observable, we use in our example the full prohibition of long positions opening (although in practice such a strict steadiness test would hardly make any sense).

The right chart of Figure 2.6.1 shows that even such a significant change of the strategy structure as complete prohibition of long positions did not change the optimization surface shape fundamentally. The global maximum shifted from the node with coordinates of 30 and 105 (at horizontal and vertical axes, respectively) to the node with coordinates of 30 and 70. This means that the optimal value of one parameter did not change at all and the optimum of the second parameter decreased by a third.

The surface of the original optimal area increased significantly. Besides, several new optimal areas emerged (all optimal areas are grouped in one part of the optimization space, around the original optimal area). These changes can be explained by the fact that during the past 10 years (the period of optimization), despite the recent financial crisis, markets were growing at moderate rates with low

volatility prevailing during most periods. In such circumstances, short option positions were more profitable than the long ones. Consequently, full prohibition of long positions led to a higher quantity of parameter combinations that turned out to be more profitable (in other words, the optimization space contains more nodes with relatively high values of objective functions).

Considering the fact that complete prohibition of long positions represents a very significant structural change (on the verge of fundamental strategy change), we should conclude that the steadiness of the optimization space in this example is rather high.

2.6.3. Steadiness Relative to the Optimization Period

Algorithmic computation of the objective function in the process of parametric optimization is based on historical data. The decision on the length of historical period (also called "historical optimization horizon") may affect optimization results significantly. Unfortunately, there are no objective criteria to select the span of historical data. Optimization should be based on data including both unfavorable and favorable periods for the strategy under development. Sticking to this concept, some automated trading system developers suggest using the database composed of separate historical frames instead of basing optimization on the continuous time series. This approach seems to be unacceptable since the choice of "appropriate" segments of historical data can hardly be objective.

In our opinion, the continuous price series should be used for parametric optimization. The decision on the length of this series should be based on the peculiarities of the specific trading strategy. The less the optimization space is sensitive to small alterations in optimization period length, the higher the steadiness of the strategy. It is desirable that the optimization space preserves its original form even under significant changes in the historical period length. The situation can be judged as ideal when the results of the optimization performed at a 2-year period are close to the results obtained at a 3-year period or when the results of 6-year and 10-year optimizations do not differ significantly.

To illustrate the idea of optimization space steadiness to changes in optimization period, we will perform the sensitivity test using the basic delta-neutral strategy and the profit-based objective function. In order to determine the extent of the steadiness of this space relative to the length of the historical period, we will perform a successive series of optimizations using historical periods of different lengths. All optimizations discussed earlier in this chapter were based on a 10-year historical period. Now we will analyze optimizations performed on 9,

8, ..., 1 years of price history.

To make optimization results comparable (and to enable the creation of the convolution of several optimizations), the values of objective functions in each case should be of approximately the same scale. It can be reached by applying the transformation described in the section dedicated to multicriteria analysis (see equation 2.4.1).

Comparison of ten optimization spaces (see Figure 2.6.2; to simplify perception of the charts, only optimal areas are shown) reveals the presence of two optimal areas in almost all surfaces. One of them is situated within 28 to 32 days by the “number of days to options expiration” parameter and 85 to 140 days by the “HV estimation period” parameter. (We will refer to this area as “the left.”) The second optimal area has coordinates of 108 to 112 by the first parameter and 90 to 160 by the second one. (This area will be referred to as “the right.”)

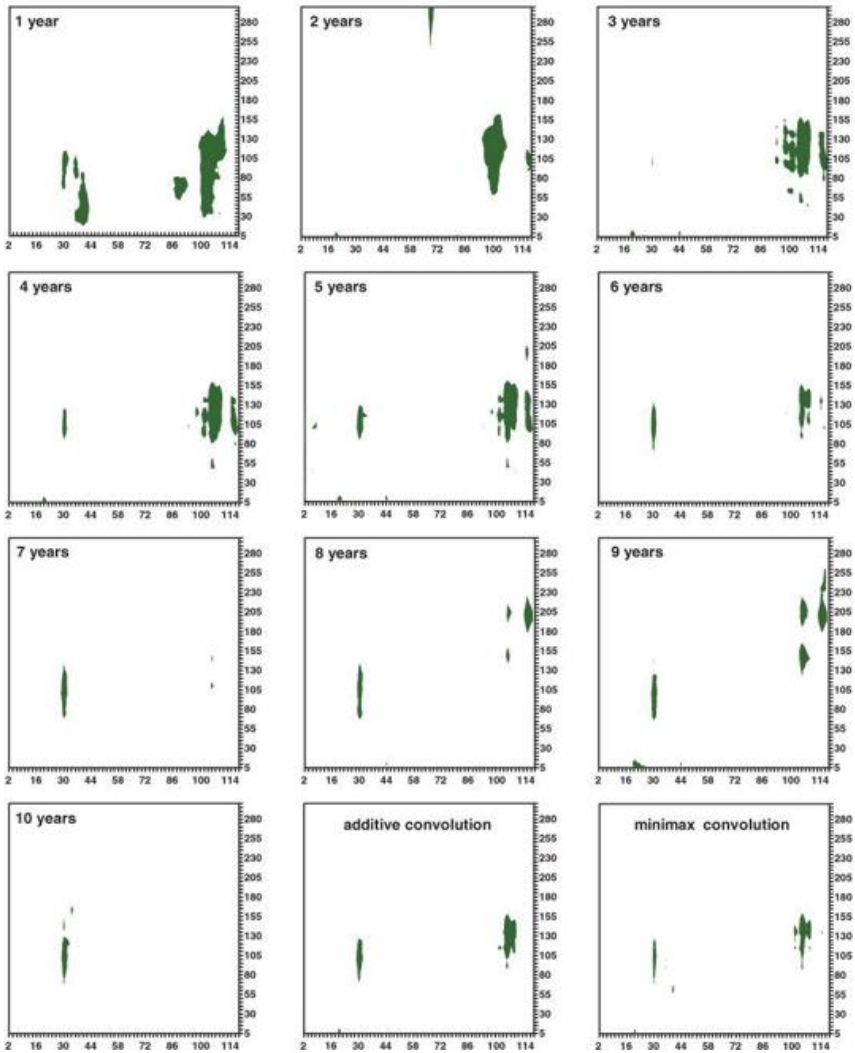

Figure 2.6.2. Variability of the optimization space of the basic delta-neutral strategy (only optimal areas are shown) induced by the change in the length of the historical periods used for strategy optimization. Each chart corresponds to one of the historical periods. The bottom-middle and bottom-right charts show minimax and additive convolution of ten optimization surfaces.

The left area is present on all surfaces except those corresponding to optimizations performed using two- and three-year time series. The area of its surface is rather stable (it changes just slightly from case to case). The only exception is the optimization based on one-year data. In this case the left area occupies a bigger space and in fact represents three separate areas grouped around the same location.

The right area is also present on all surfaces, excluding the only case when optimization was performed using a ten-year historical period. In two more cases, this area turned out to be slightly shifted toward higher values of the “HV estimation period” parameter. The size of the right area is more variable than that of the left area. The right area of optimizations based on one-, two-, three-, four-, and five-year data is of a rather large size. In contrast to that, on optimization surfaces obtained at longer periods it has a smaller size. Besides, in all cases the right area does not form a single whole but is rather divided into several sub-areas.

The extent of optimization space steadiness can be estimated by different methods. The simplest of them consists in visual analysis. The comparison of different charts in Figure 2.6.2 suggests that this optimization is rather steady. This conclusion follows from the above-described persistent disposition of optimal areas within the optimization space. More sophisticated approaches to estimating the steadiness should express numerically the sensitivity of optimization space (for example, by calculating the variability of coordinates of nodes constituting the optimal areas).

When steadiness is estimated by comparing the large number of optimization spaces (as it is in our example), the convolution method can be used to facilitate the comparison. If spaces differ from one another considerably (that is, optimization is unsteady), their convolution will have an appearance of a surface with a large number of optimal areas scattered over it irregularly. At the same time, if optimization spaces are similar, with optimal areas having approximately the same form and situated at the more or less same locations (indicating steadiness of optimization), the convolution will have a limited number of easily distinguishable optimal areas. The bottom-middle and bottom-right charts of Figure 2.6.2 show additive and minimax convolutions of ten optimization surfaces (both convolutions produce in this case almost identical results). The left and the right optimal areas are localized clearly and can be distinguished easily. This confirms once again our conclusion about optimization steadiness relative to the change of the historical period used for strategy optimization.

2.7. Optimization Methods

Hitherto, we used the most informative optimization method—exhaustive search through all possible combinations of parameter values. This method requires estimating the values of objective function(s) in all nodes of the optimization space. Since the computation of the objective function in each node requires an

execution of complex multistep procedures, the main drawback of an exhaustive search is the long time required to complete the full optimization cycle. As the number of parameters increases, the number of computations and the time increase according to the power law. The expansion of the acceptable value range and reduction of the optimization step also increase the time required for an exhaustive search.

In all examples considered up to this point, only two parameters were optimized and 60 values were tested for each of them. Accordingly, optimization space consisted of 3,600 nodes. On average, one computation (calculation of one objective function in one node) took about one minute (using the system of parallel calculations running on several modern PCs). This means that construction of full optimization space similar to those considered earlier takes about 60 hours. Furthermore, adding just one extra parameter to the optimization system (three in total) increases calculation time dramatically, up to five months! Obviously, this is unacceptable for efficient creation and modification of automated trading strategies. This implies that an exhaustive search is not the most efficient (and in most cases not suitable at all) method to search for optimal solutions.

In most cases, it would be more reasonable to apply special techniques that avoid an exhaustive search through all the nodes of the optimization space. A whole group of methods—from the simplest to the most complex ones—can be used to maximize (or to approach the maximum) the objective function via *direct (as opposed to the exhaustive) search*. This area of applied mathematics has recently achieved considerable success and continues to develop rapidly. In this review, we will not touch on the two most popular optimization methods—genetic algorithms and neuronets. Firstly, these methods are so complex that each of them would require a separate book (the number of publications on this topic grows from year to year). Secondly, most challenges of parametric optimization with a limited number of parameters can be solved at a lower cost. In this section, we consider five optimization methods that were selected on the basis of their effectiveness (considering the specific problems encountered in optimization of options trading systems) and simplicity of practical implementation.

Optimization methods that do not require an exhaustive search have two significant drawbacks. They reach maximal efficiency only when the optimization space is unimodal (that is, contains a single global maximum). If there are several local extremes (when optimization space is polymodal), these methods may get to a sub-optimal solution (that is, select the local extreme instead of the global maximum).

Figuratively speaking, this situation may be compared with climbing a hill instead of capturing a mountain peak nearby. This is a general drawback of all methods that are based on the principle of gradually improving the value of objective function by moving around the node selected at the previous step of the optimization algorithm. This is an unavoidable drawback inherent in all optimization methods that rely on direct search. As we noted earlier, finding the global maximum can be guaranteed only by an exhaustive search through all the optimization space nodes.

There is a simple, although time-consuming, way to solve this problem. If there are reasons to believe that the optimization space contains more than just one extreme, optimization should be run repeatedly with different starting conditions (that is, each time it begins from a different node). Starting nodes may be allocated evenly within the optimization space, or they may be selected randomly. Although reaching the global maximum is still not guaranteed, the probability of finding it can be brought up to an acceptable level. Certainly, the more starting points that are employed, the higher the probability is that at least one of them will lead to the highest extreme (however, time costs also increase rapidly). The broader the optimization space is, the more starting points are needed.

The second drawback is that the solution selected by applying one of the direct search methods does not contain the information about objective function values in nodes adjacent to the node of the optimal solution. Consequently, the developer is unable to determine the properties of the optimal area surrounding the extreme that has been found. Hence, the robustness of the optimal solution cannot be estimated. As we repeatedly outlined earlier, robustness is one of the main properties that defines optimization reliability. Inability to assess the robustness may challenge the validity of optimal solution. However, this problem can be solved by calculating objective function values in all nodes surrounding the optimal solution. After that the robustness can be estimated using the methods described in section 2.5.

2.7.1. A Review of the Key Direct Search Methods

Alternating-Variable Ascent Method

The essence of the alternating-variable ascent method (the word “descent” is usually used in the name of this method, but, as noted earlier, it is preferable to solve the problem of profit maximization in optimization of trading strategies) consists in a successive search by each parameter when they are taken in turn. The algorithm of this method is the following:

1. Select the starting node from which the optimization process begins (the selection of the starting point is required for initiation of any direct search method). The selection may be random, deliberate (that is, based on prior knowledge of the developer), or predetermined (for example, if many iterative optimizations are anticipated, starting nodes may be distributed in the optimization space evenly).

2. Since the parameters are optimized successively rather than simultaneously, the order should be determined. In most cases the order is not of big importance. However, if one parameter is more important than the others, the optimization should start with it.

3. Beginning from the starting point, the best solution is found for the first parameter (values of all other parameters are fixed at the starting node values). The search is performed using any one-dimensional optimization method. If the number of nodes to be calculated is relatively low, we can even use the exhaustive search method.

4. When the best solution for the first parameter is found, the algorithm passes on to the optimization of the next parameter. The one-dimensional optimization is performed once again; values of all other parameters are fixed at values of the node that has been found during optimization of the first parameter.

5. When one optimization cycle (successive one-dimensional optimization of all parameters) is completed, the procedure is reiterated starting from the last node. The process stops when the next optimization cycle does not find the solution that prevails over the solution found during the preceding cycle.

We will demonstrate the practical application of this algorithm by considering an example of optimization of basic delta-neutral strategy. The optimization space of this strategy obtained using the profit-based objective function is shown in Figure 2.2.2. To present the search procedure visually, we will limit the permissible ranges of parameter values to 2 to 80 for the “number of days to options expiration” parameter and 100 to 300 for the “HV estimation period” parameter. The algorithm is executed in the following way:

1. The starting point is selected. Suppose that the node with coordinates of 60 and 130 was selected randomly. This node is marked with number 1 in Figure 2.7.1.

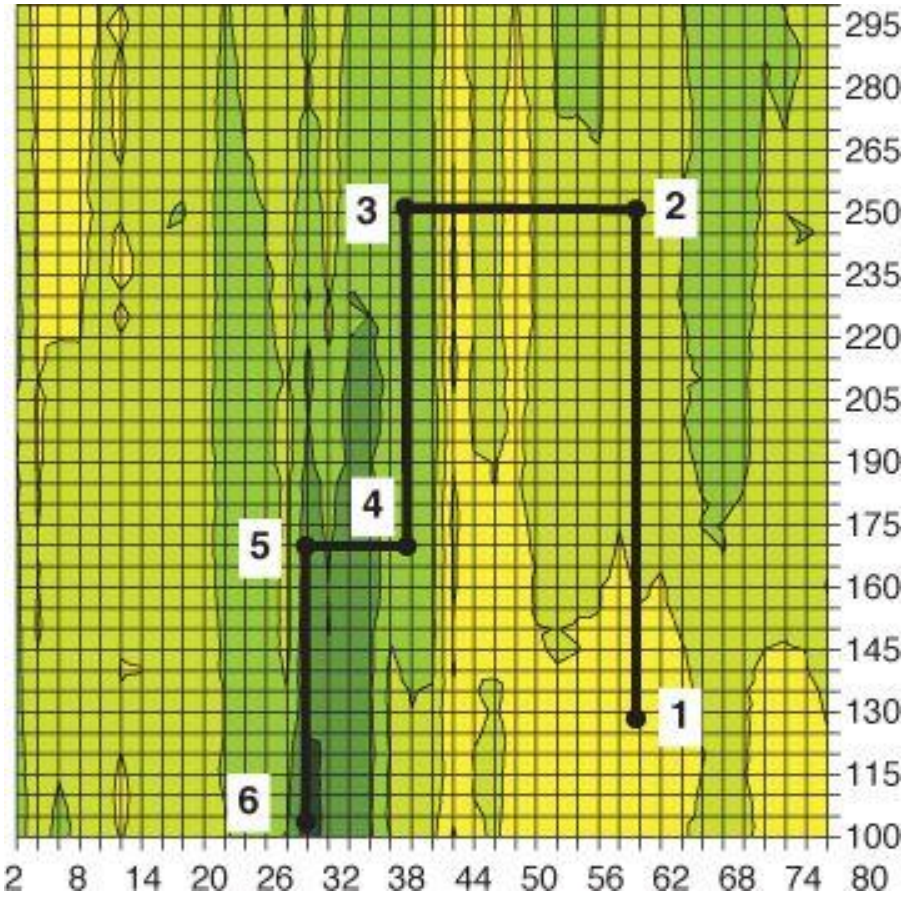

Figure 2.7.1. Graphical representation of optimization performed using the alternating-variable ascent method.

2. With the "number of days to options expiration" parameter fixed at 60, we compute the objective function for all values of the "HV estimation period" parameter (one-dimensional optimization using an exhaustive search).

3. The node with the maximum objective function value is found. In this case it is the node with coordinates of 60 and 250 (point number 2 in Figure 2.7.1).

4. The value of the "HV estimation period" parameter is fixed at 250, and the objective function for all values of the "number of days to options expiration" parameter is calculated. The objective function is maximal at the node with coordinates of 40 and 250 (third point).

5. The number of days to options expiration is fixed at 40 and the objective function for all values of the HV estimation period is calculated. In this way we get to the next node with

coordinates of 40 and 170 (fourth point).

6. The HV estimation period is fixed at 170 and the objective function for all days to options expiration is calculated. The fifth point with coordinates of 30 and 170 is, thereby, obtained.

7. The number of days to expiration is fixed at 30 and the objective function is calculated for all HV estimation periods. The sixth point has coordinates of 30 and 105.

8. The HV estimation period is fixed at 105 and the objective function is calculated for all days to options expiration. It turns out that the objective function does not improve any further and, hence, the algorithm stops. The last node (number 6) represents the optimal solution.

The alternating-variable ascent method, despite being simple in its logic and easy to implement, is not very efficient. Imagine the situation when two-dimensional optimization surface has the optimal area in the form of a narrow ridge stretching from "northwest" to "southeast." If the height of the ridge increases toward either the southeast or the northwest, this algorithm will climb to the crest of the ridge and it would be impossible to improve this position since two parameters must be amended simultaneously for that. This restricts heavily the applicability of the alternating-variable ascent method.

Hook-Jeeves Method

The Hook-Jeeves method was developed to avoid situations in which the alternating-variable ascent method stops prematurely without finding the satisfactory optimal solution. It consists in consecutive execution of two procedures—exploratory search and pattern-matching search reiterated until the unimprovable optimum is found. As compared to the alternating-variable ascent algorithm, the Hook-Jeeves method significantly lowers the possibility of the algorithm stopping without finding the extreme. The sequence of actions for implementing the Hook-Jeeves method is as follows:

1. Select the starting node. Beginning from this node, the exploratory search procedure is performed. This procedure consists in execution of several cycles of the alternating-variable ascent method. One cycle of exploratory search consists of n cycles of alternating-variable search (n is the number of parameters). For two-dimensional optimization space, there are two cycles of alternating-variable ascent that result in finding sub-optimal nodes in addition to the starting node.

2. Execute the pattern search procedure. For this, determine the direction from the starting node to the node found as a result of the first cycle of exploratory search. The value of the objective function increases in this direction, and it would be reasonable to assume that moving farther in this direction will bring about further improvement.

3. Continue the search in the direction determined in the preceding stage. In contrast to the alternating-variable ascent method, this search is performed under the simultaneous changes in all parameters. Whereas in the alternating-variable ascent method the search is allowed only in the horizontal or the vertical directions (in the two-dimensional case), the pattern-matching search allows moving in any direction within the optimization space.

4. Having found the best node in this direction, the cycle of the exploratory search is repeated, after which the pattern search is executed again. These altering cycles continue until the node that cannot be improved any further is found.

Let us consider the application of the Hook-Jeeves method using basic delta-neutral strategy as an example (optimization is performed by applying the profit-based objective function). Similar to the previous example, the acceptable values of parameters will be limited to the same ranges (2 to 80 for “number of days to options expiration” and 100 to 300 for “HV estimation period”). The algorithm is executed in the following way:

1. The starting point is selected randomly. Suppose that this is the node with coordinates of 68 and 300 (marked with number 1 in Figure 2.7.2).

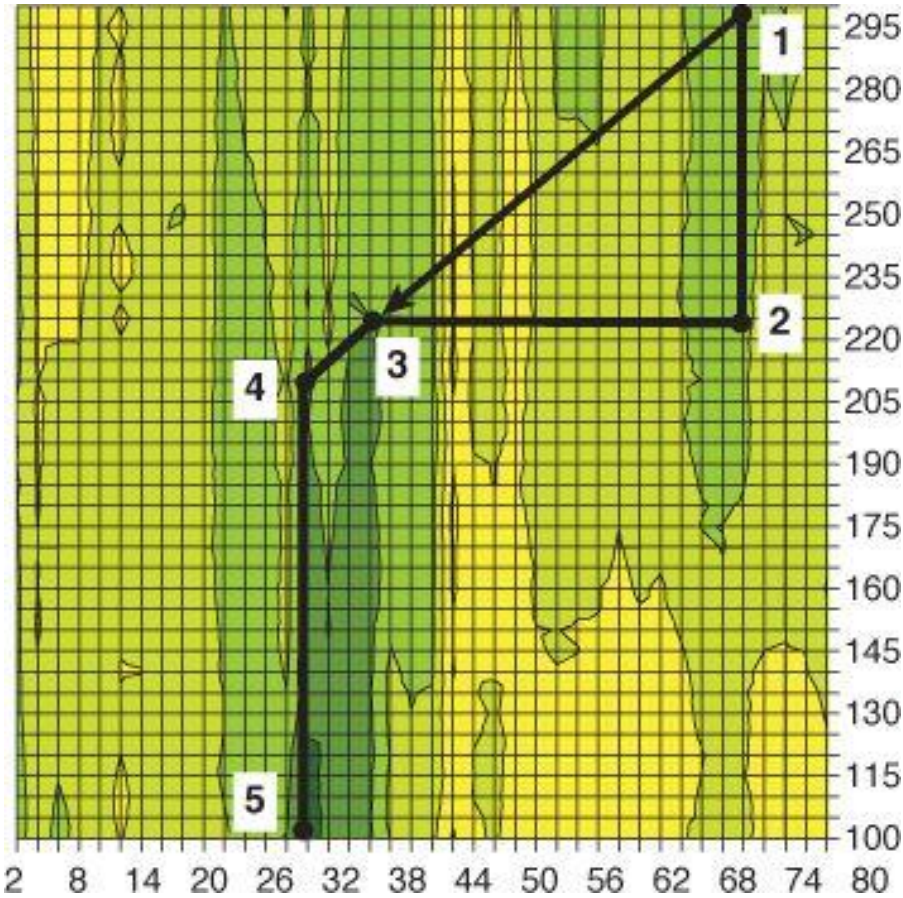

Figure 2.7.2. Graphical representation of optimization performed using the Hook-Jeeves method.

2. The procedure of exploratory search consists of two cycles (equal to the number of parameters) of alternating-variable ascent method. Having the "number of days to options expiration" parameter fixed at 68, the objective function is calculated for all values of the "HV estimation period" parameter. The maximum value of the objective function falls at the node with coordinates 30 and 225 (the second point in Figure 2.7.2).

3. The value of "HV estimation period" is fixed at 225, and the objective function is calculated for all values of the "number of days to options expiration" parameter. The maximum function value corresponds to the node with coordinates of 36 and 225 (the third point).

4. The pattern-search procedure is performed. The direction from the starting node to the third node is determined (the

arrow in Figure 2.7.2), and a one-dimensional optimization procedure is executed in this direction. The values of the objective function are calculated at all nodes crossed by the line drawn in the given direction.

5. The node with the maximum objective function value (the fourth point with coordinates of 30 and 210) is selected. Starting from this node, the exploratory search cycle is repeated.

6. The "number of days to options expiration" parameter is fixed at 30, and the objective function is calculated for all values of the "HV estimation period" parameter. The maximum function value falls at the node with coordinates of 30 and 105 (the fifth point).

7. The "HV estimation period" is fixed at 105, and the objective function is calculated for all days to options expiration. Since there is no further improvement for the objective function, the algorithm stops. The node number 5 is selected as the optimal solution.

Rosenbrock Method

The Rosenbrock method, also called "the method of rotating coordinates," is a step further in the elaboration of the alternating-variable ascent method. The first stage of the Rosenbrock algorithm coincides with the alternating-variable ascent method. Then, similar to the Hook-Jeeves method, we find the direction in which the objective function improves. However, apart from this direction, we create (n - 1) more directions, each of which is orthogonal to the found direction and all of which are orthogonal to each other. Thus, instead of the original coordinate system, defined by the parameters, a new coordinate system is established. Relative to the initial coordinate system, it is rotated so that one of its axes coincides with the direction in which the objective function improves. In the new coordinate system, we perform the next alternating-variable ascent cycle, the result of which is used for determining the new improvement direction followed by the new rotation of the coordinate system. For a two-dimensional optimization space the Rosenbrock algorithm can be presented in the following way:

1. Select the starting node where the cycle of exploratory search (that is, two cycles of alternating-variable ascent search) begins.

2. Determine the direction from the starting node to the node found at the stage of exploratory search, and perform the

procedure of pattern-matching search in this direction.

3. Having found the best node out of those situated on the line drawn in this direction, build the new direction orthogonal to the original one.

4. Using the alternating-variable ascent method, perform the search in the orthogonal direction and find the node with the highest objective function value.

5. Repeat all procedures beginning from the node found in the orthogonal direction.

Next we illustrate the application of the Rosenbrock method to optimization of the basic delta-neutral strategy. The objective function and the limits imposed on acceptable ranges of parameter values are the same as in previous examples. The algorithm is executed in the following way:

1. Select the starting point randomly. Suppose that the node with coordinates of 72 and 180 was chosen (point number 1 in Figure 2.7.3).

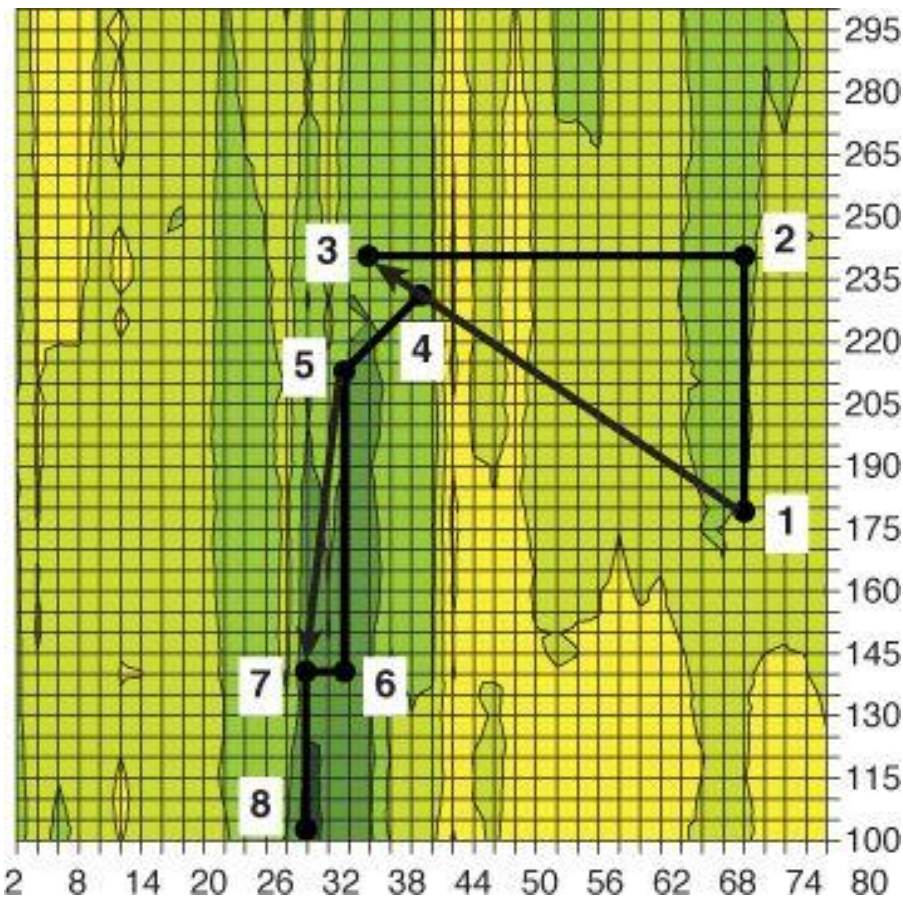

Figure 2.7.3. Graphical representation of optimization performed using the Rosenbrock method.

2. Begin the exploratory search procedure, which consists of two cycles of the alternating-variable ascent. The "number of days to options expiration" parameter is fixed at 72 and the objective function is calculated for all values of the "HV estimation period" parameter. The node with the maximum objective function value has coordinates of 72 and 240 (point number 2 in Figure 2.7.3).

3. Having fixed the value of the "HV estimation period" parameter at 240, the objective function is calculated for all values of the "number of days to options expiration" parameter. The maximum of the function falls at the node with coordinates of 36 and 240 (point number 3).

4. The pattern-search procedure is executed. The direction from the starting node to the third node is determined (the top-right arrow in Figure 2.7.3), and values of the objective function are calculated for all nodes crossed by this direction.

5. The node with the maximum objective function value is found (point number 4 with coordinates of 40 and 230), and the direction orthogonal to the one found earlier is established.

6. The search in the new (orthogonal) direction is performed using the alternating-variable ascent method, and the node with the highest objective function value (point number 5 with coordinates 34 and 215) is selected.

7. Starting from this node, the exploratory search cycle is repeated. The "number of days to options expiration" parameter is fixed at a value of 34, and the objective function for all values of the "HV estimation period" parameter is calculated. The maximum function value turns out to be at the node with coordinates 34 and 140 (point number 6).

8. The "HV estimation period" is fixed at 140 and the objective function is calculated for all days to options expiration. The function's maximum falls at the node with coordinates 30 and 140 (point number 7).

9. The pattern-search procedure is repeated. The direction from the fifth point to the seventh point is established (the bottom-left arrow in Figure 2.7.3) and the values of the objective function are calculated at all nodes crossed by this direction. It turns out that there are no nodes in this direction

for which the objective function value would exceed the value possessed by the seventh node.

10. The direction orthogonal to the one shown by the left arrow is established (not shown on the Figure), and the search is performed along this direction. It turns out that there is also no node for which the objective function value would exceed the seventh node's value.

11. The cycle of exploratory search is repeated. The value of the "number of days to options expiration" parameter is fixed at 30, and the objective function is calculated for all values of the "HV estimation period" parameter. The function's maximum falls at the node with coordinates 30 and 105 (point number 8).

12. As this has already been demonstrated in the previous examples (the Hook-Jeeves method), no further improvement is possible at this point, and hence, the algorithm stops. The eighth node is selected as the optimal solution.

The Rosenbrock method is a refinement of the alternating-variable ascent method and the Hook-Jeeves method. In some cases it can significantly improve the search effectiveness, though it is not always the case. Under certain circumstances, depending on the shape and the structure of the optimization space, the search effectiveness may even decrease due to the application of this method.

Nelder-Mead Method

The Nelder-Mead method (also called the deformable polyhedron method, simplex search, or amoeba search method) has two modifications: original variant (based on rectilinear simplex) and advanced modification (which utilizes deformable simplex). In this section we will discuss the advanced modification. Using the term "simplex" may be misleading to some extent since there is a widely known simplex method of linear programming developed to solve the problem of optimization with linear objective function and linear constraints that has nothing to do with the described method. In the Nelder-Mead method the simplex represents the polyhedron in n-dimensional space with $n+1$ vertices. It can be considered as a generalization of a triangle in the multidimensional space. The triangle is an example of the simplex in two-dimensional space.

Figuratively speaking, during optimization the simplex rolls over the parameters space, gradually approaching the extreme. Having calculated the objective function values in the vertices of the simplex, we find the worst of them and move the simplex so that all other vertices stay at their places and the vertex with the worst value is substituted with the vertex that is symmetrical to the worst one

relative to the simplex center. This can be imagined visually in the two-dimensional case when the simplex (represented by the triangle) rolls over its side that is opposite to the worst apex. By repeating such cycles, the simplex approaches the extreme until the objective function is no longer improved. The best of the obtained apexes is accepted as the optimal solution. When the simplex is in the immediate vicinity of the extreme, so that the distance from its center to the extreme is less than the simplex side, it loses the capability to approach the extreme any closer. In this case we can decrease the simplex size, while maintaining its initial shape and continue our search by repeating the size decrease each time when we lose the capability to get closer to the extreme, until the side length becomes smaller than the optimization step.

Besides rolling over the optimization space, the deformable simplex may change its shape (which explains another term—"amoeba method"). While the method of rectilinear simplex has no other parameters except the length of the simplex edge, the deformable simplex method involves the system of four parameters: reflection coefficient α, expansion coefficient σ, contraction coefficient γ, reduction coefficient ρ. As experience has shown, the choice of coefficient values may be critical for obtaining satisfactory optimization results.

The algorithm of the Nelder-Mead method consists of the following steps:

1. Select $n+1$ points of the initial simplex x_1, x_2... x_{n+1}. In the two-dimensional case, when the simplex represents the triangle, it is sufficient that three points are not located on the same straight line.

2. Range points by the objective function values (given that the function has to be maximized):

$$f(x_1) \geq f(x_2) \geq f(x_3) \geq \ldots \geq f(x_{n+1})$$

3. Calculate point x_0, which is the center of the figure whose vertices coincide with simplex vertices, except the worst one x_{n+1}:

$$x_0 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

For two-dimensional optimization x0 is located in the middle of the segment connecting the best and intermediate vertices (as estimated by the values of the objective function) of the triangle.

4. Reflection step. Calculate the reflected point:

$$x_r = x_0 + \alpha(x_0 - x_{n+1}).$$

If the objective function value at the reflected point is better than at the second-worst point x_n, but is not better than at the best point x_1, then the reflected point substitutes in the simplex the excluded worst point and the algorithm returns to step 2 (simplex rolls over from the worst point toward the point with a higher objective function value). If the objective function value at the reflected point is better than at all initial points, the algorithm moves to step 5.

5. Expansion step. Calculate the expansion point. It is located along the line passing through the worst simplex point and the simplex center. Its position is determined by the expansion coefficient:

$$x_e = x_0 + \sigma(x_0 - x_{n+1}).$$

The essence of this operation is the following: If the direction in which the objective function increases is established, the simplex should be expanded in this direction. If the expansion point is better than the reflection point, the latter is substituted in the simplex by this expansion point and the simplex becomes expanded in this direction. After that the algorithm returns to step 2. If the expansion point is not better than the reflected point, there is no sense in expansion, the reflected point is left in the simplex and its shape does not change. The algorithm also returns to step 2.

6. Contraction step. The algorithm gets to this step if the reflected point is not better than at least the second-worst point. The contraction point is calculated as

$$x_c = x_0 + \rho(x_0 - x_{n+1}).$$

If the contraction point is better than the worst point, the algorithm substitutes the former for the latter and passes on to step 2. As a result, the simplex contracts in this direction. If the obtained point is worse than the worst point, it may testify to the fact that we are already in the immediate vicinity of the extreme, and to find it the simplex size has to be decreased. The algorithm proceeds to step 7.

7. Reduction step. The point with the best objective function value keeps its position, while all other points are pulled up

toward it.

$$x_i = x_0 + \gamma(x_i - x_1),\ \forall i \in \{2,\ldots n+1\}.$$

If the simplex size (considering its irregular shape, we need to define it specially, for example, as the average distance from the center to all vertices) turns out to be less than the specified value, the algorithm stops. Otherwise, it goes to step 2.

Let us consider the practical application of the Nelder-Mead method for optimization of the basic delta-neutral strategy. The objective function and the limits imposed on acceptable ranges of parameter values are the same as in previous examples. The algorithm is executed in the following way:

1. Select three points of the initial simplex. Assume that nodes with coordinates 12 and 105, 18 and 110, 18 and 100 are selected as simplex vertexes.

2. Calculate the values of the objective function for the three points of the initial simplex and find the node with the worst function value. This node is marked with number 1 in Figure 2.7.4.

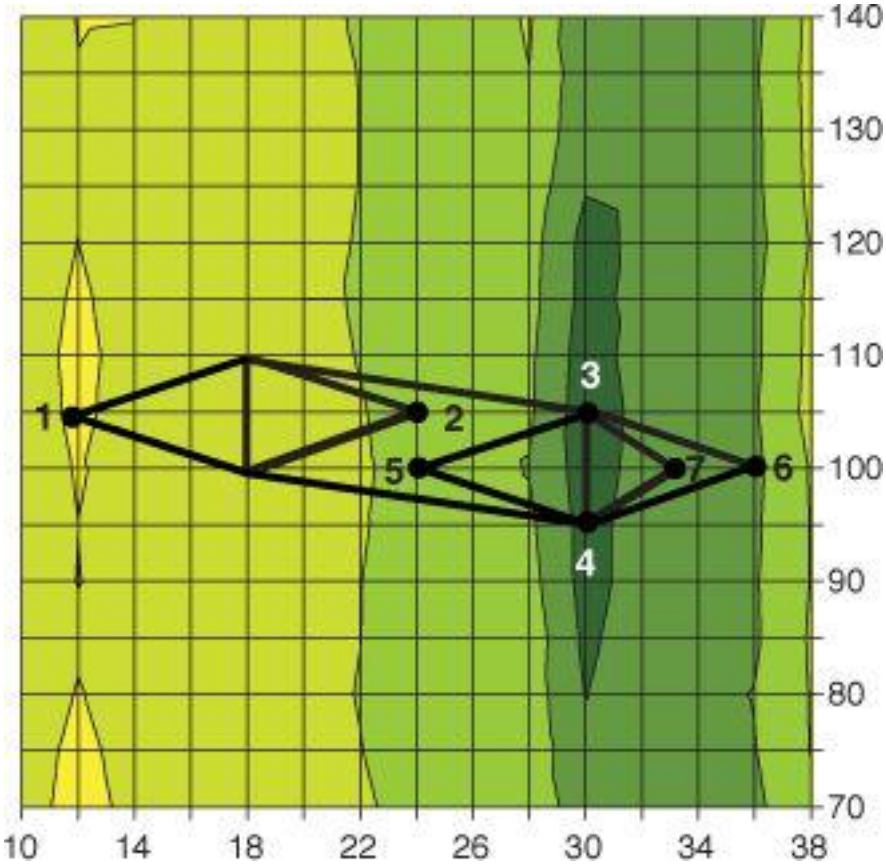

Figure 2.7.4. Graphical representation of optimization performed using the Nelder-Mead method.

3. Calculate the center of the segment with coordinates 18 and 110, 18 and 100. The center is the node with coordinates 18 and 105.

4. Execute the reflection step using the reflection coefficient $\alpha = 1$. The reflected point (marked with number 2 in Figure 2.7.4) has coordinates of 24 and 105. Since the objective function value in the reflected point is better than in all points of the initial simplex, the algorithm passes on to the expansion step.

5. Calculate the expansion point using the expansion coefficient $\sigma = 2$. It is found by doubling the distance between the simplex center and the second point. The node corresponding to the expansion point (number 3 in Figure 2.7.4) has a higher objective function value as compared to node number 2 and hence substitutes for the latter becoming the new vertex of the simplex.

6. Execute another reflection step. The reflected point corresponds to point number 4 in Figure 2.7.4. Since the objective function value in the reflected point is lower than in the second-worst point of the previous simplex, the algorithm passes on to the contraction step.

7. Calculate the contraction point using the contraction coefficient $\gamma = 0.5$ (this is node number 5). Since this node is better than the worst one in the previous simplex, but is not the best, the algorithm passes on to the reflection step.

8. Having executed the reflection step, node number 6 is obtained. Since the objective function value in this node is lower than in all points of the previous simplex, the algorithm passes on to the contraction step.

9. Having calculated the contraction point, point number 7 is obtained. Since this point falls between two nodes, further execution of standard Nelder-Mead algorithm for this optimization space is impossible. To select the final optimal solution, we can go several ways. The objective function of point 7 can be calculated by interpolation as the mean of objective functions of adjacent nodes (with coordinates of 32 and 100, 34 and 100). Then the standard algorithm can be resumed (by treating all points falling outside the nodes in the same way). Another, simpler approach consists in selecting the optimal solution among vertices of the last simplex (the one having the highest objective function value).

This algorithm proves to be rather effective in solving different kinds of optimization problems. Its main drawback is a large number of coefficients, the choice among which may affect optimization effectiveness significantly. As shown in the next section, accurate selection of coefficient values and preliminary information about optimization space properties are necessary to apply the Nelder-Mead method effectively.

2.7.2. Comparison of the Effectiveness of Direct Search Methods

In the preceding section we described four direct search methods. Each of them has its specific features, advantages, and drawbacks. The first method (the alternating-variable ascent method) is the most basic and the simplest. All others are more complex and each one is more complex and more sophisticated than the preceding ones (according to the order of their description in section 2.7.1). However, our experience shows that more complex methods are not necessarily more effective. In general, we can state that the effectiveness of any

particular method depends on the shape of the optimization space to which it applies. A method that has been proved to be quite effective for a given type of optimization space may be inappropriate for a space with a different structure.

In this section we test effectiveness of the four methods described earlier and analyze the dependence of their effectiveness on the shape of the optimization space. For that purpose, each of the four search methods will be applied to two optimization spaces (corresponding to the basic delta-neutral strategy) with different shapes. One of the spaces has a single optimal area of a relatively large size (it is formed by applying the profit-based objective function; see the top-left chart of Figure 2.3.1). The second space (formed by applying the objective function that is based on the percentage of profitable trades; see the bottom-left chart of Figure 2.3.1) has a totally different shape—a multitude of small optimal areas.

To obtain statistically reliable results, we executed 300 full optimization cycles for each method and for each of the two objective functions. The cycles differed from each other only by the initial point where the search for optimal solution starts. The comparative analysis of effectiveness of different optimization methods will be based on the following five characteristics (in each case they are calculated on the basis of 300 input data):

- **Percentage of hitting the global maximum**—The percentage of optimal solutions coinciding with the global maximum (when the optimization algorithm stops at the node with the highest objective function value).
- **Percentage of hitting the optimal area**—The percentage of optimal solutions falling within the optimal area. The objective function values of these solutions are higher than or equal to the predetermined threshold level.
- **Percentage of hitting the poor areas**—The percentage of unsatisfactory optimal solutions. The objective function values of these solutions are not exceeding some predetermined threshold level.
- **Average value of optimal solutions**—Average of the objective function values of all optimal solutions. The average is calculated using the data of 300 optimization cycles.
- **Average number of computations**—Average number of calculations necessary to execute one complete optimization cycle. The average is based on 300 optimization cycles.

Table 2.7.1 shows values of five effectiveness characteristics

corresponding to four search methods. When the objective function is based on profit, the Hook-Jeeves method demonstrates the highest effectiveness. In 69% of 300 optimization cycles, the node corresponding to the global maximum was selected as the optimal solution (accordingly, in 31% of cases the algorithm stopped without finding the node with the highest value of the objective function). This method also considerably surpassed other methods by the percentage of hitting the optimal area (72% of cases) and average value of the optimal solution (6.43). The only ratio by which the Hook-Jeeves method is slightly worse than the alternating-variable ascent method is the percentage of hitting the poor areas (15% and 14%, respectively). However, the significant drawback of this method is that it is highly time-consuming. One optimization cycle requires 1,279 calculations on average, which is much more than the number of calculations needed for other methods.

Table 2.7.1. Effectiveness characteristics of four optimization methods based on the direct search of optimal solutions.

Objective Function: Profit					
Optimization Method	**Percentage of Hitting the Global Maximum (7.13)**	**Percentage of Hitting the Optimal Area (>6)**	**Percentage of Hitting the Poor Areas (<5)**	**Average Value of Optimal Solutions**	**Average Number of Computations**
Alternating-variable ascent method	48	48	14	6.15	319
Hook-Jeeves method	69	72	15	6.43	1,279
Rosenbrock method	11	29	41	5.21	835
Nelder-Mead method	7	15	60	4.45	107
Objective Function: Percentage of Profitable Trades					
Optimization Method	**Percentage of Hitting the Global Maximum (0.74)**	**Percentage of Hitting the Optimal Area (>0.7)**	**Percentage of Hitting the Poor Areas (<0.65)**	**Average Value of Optimal Solutions**	**Average Number of Computations**
Alternating-variable ascent method	8	37	0	0.69	226
Hook-Jeeves method	9	38	0	0.69	959
Rosenbrock method	12	28	1	0.69	712
Nelder-Mead method	2	2	56	0.64	106

The alternating-variable ascent method takes second place by the effectiveness of optimizing the unimodal profit-based objective function. Although it is inferior to the Hook-Jeeves method by all indicators (except the percentage of hitting the poor areas), its indisputable advantage is the relatively low number of calculations required to complete one optimization cycle (319 on average, which is one-fourth of the number for the Hook-Jeeves method).

The effectiveness of the Rosenbrock method is inferior to the two previous methods. Only in 11% of the 300 trials did this method manage to discover the global maximum, and in 41% of the trials the algorithm stopped at nodes situated in areas characterized by low values of the objective function. Besides, the number of calculations required to find optimal solutions using the Rosenbrock method turned out to be quite high—835 on average, which is slightly better than the value obtained for the Hook-Jeeves method and much worse than the figures relating to other methods.

The worst results were demonstrated by the Nelder-Mead method. Although the number of calculations required by this algorithm was superior to other methods, all indicators characterizing its effectiveness have very low values. In only 7% of all trials, the optimal solutions found by this method coincided with the global maximum; and in only 15% of cases, the optimal solution was located in the optimal area. The solutions were unsatisfactory in as much as 60% of the trials.

When applied to another optimization space (constructed using the objective function based on the percentage of profitable trades), the effectiveness rates of the alternating-variable ascent method, the Hook-Jeeves method, and the Rosenbrock method were approximately the same (except for the fact that the last method was slightly worse by the percentage of hitting the optimal area). Again, the Nelder-Mead method showed the worst results. Such low effectiveness of this algorithm is quite surprising. Perhaps, the reason is that the initial conditions (the simplex size) and the values of its numerous coefficients (reflection, contraction, reduction, and expansion) should be thoroughly selected to implement this method effectively. We selected the simplex size arbitrarily and applied commonly used values for the coefficients. To obtain satisfactory results, we should probably define the coefficients more deliberately and select the initial simplex on the basis of some a priori assumptions about the optimization space properties.

The effectiveness of all four optimization methods was lower for the optimization space corresponding to the objective function based on the percentage of profitable trades (as compared to the space

constructed using the profit-based function). This finding supports our supposition that the shape of the optimization space influences the effectiveness of optimization greatly. Most likely, *unimodal spaces with a relatively broad optimal area are easier to optimize by direct search methods than polymodal spaces with a large number of scattered optimal areas*.

2.7.3. Random Search

Up to here we discussed two approaches to optimization—exhaustive search, requiring calculation of the objective function in all optimization space nodes, and direct search. There is another approach consisting of random selection of optimization space nodes. Certainly, random search is the simplest optimization method. To implement it, one has to select the specified number of nodes randomly and calculate their objective function values. Then the node with the maximum value is selected as the optimal solution. This method is so primitive that it is often not even considered as an appropriate alternative. Nevertheless, in many cases (and by many characteristics) random search may ensure rather effective results not inferior to direct search methods.

The main and the only factor affecting the effectiveness of the random search is the number of nodes to be selected. The more trials that are made, the more probable it is that the optimal solution will coincide with the global maximum or lie in the closest vicinity to it. The number of trials should depend on the size of the optimization space. In all our examples optimization space consists of 3,600 nodes. If we execute only 100 trials in such a space, less than 3% of the nodes will be tested, which is apparently insufficient to select a satisfactory optimal solution. However, if optimization space consists of only 500 nodes, 100 trials cover 20% of the space, which may be quite sufficient.

In this section we will analyze three aspects of random search effectiveness:

1. The relationship between search effectiveness and the number of randomly selected nodes. The effectiveness will be tested for 100, 200, ..., 1,000 trials (in total, 10 variants of the selected node's quantities).

2. The influence of the optimization space shape on search effectiveness. As in the preceding section, the random search method will be applied to two different optimization spaces —one constructed using the profit-based objective function and another one relating to the function based on the percentage of profitable trades.

3. Comparison of random search effectiveness with the effectiveness of two direct methods—the alternating-variable ascent method and the Hook-Jeeves method (in the preceding section these algorithms were proved to be the most effective among other direct search methods).

To analyze the effectiveness of the random search, we will use the same characteristics as for comparing four direct search methods (see Table 2.7.1).

As one would expect, values of all characteristics increase as the number of randomly selected nodes rises (see Figures 2.7.5 and 2.7.6). However, the rates of these increases vary for different characteristics. Moreover, the shape of the relationship between search effectiveness and the number of trials is also characteristic-dependent. The percentage of optimal solutions coinciding with the global maximum increases linearly as the number of selected nodes increases. On the other hand, the percentage of optimal solutions located in the optimal area and the average value of objective function increase nonlinearly as the number of trials increases: Initially the growth in values of these characteristics is rapid, but a further increase in the number of trials adds only slightly to the search effectiveness.

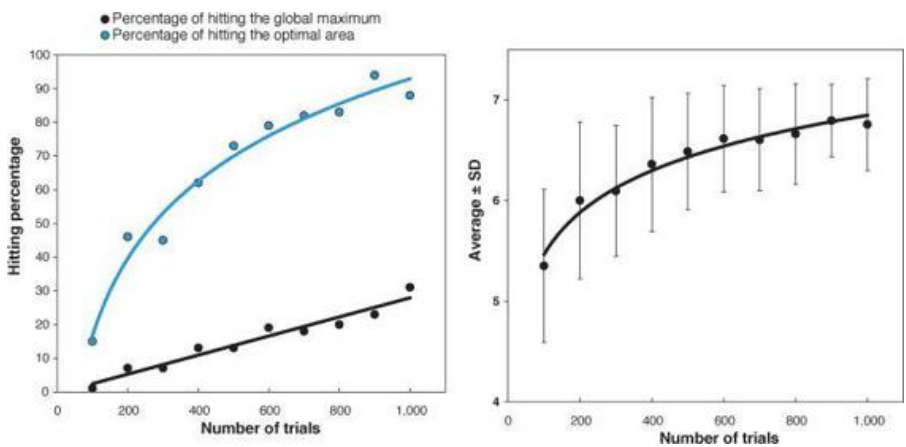

Figure 2.7.5. The relationship between the effectiveness of random search and the number of selected nodes (trials). Objective function: profit. The left-hand chart shows the relationship for two characteristics: percentage of hitting the global maximum and percentage of hitting the optimal area. The right-hand chart shows the average value of the optimal solution (another effectiveness characteristic) and its variability in the form of standard error.

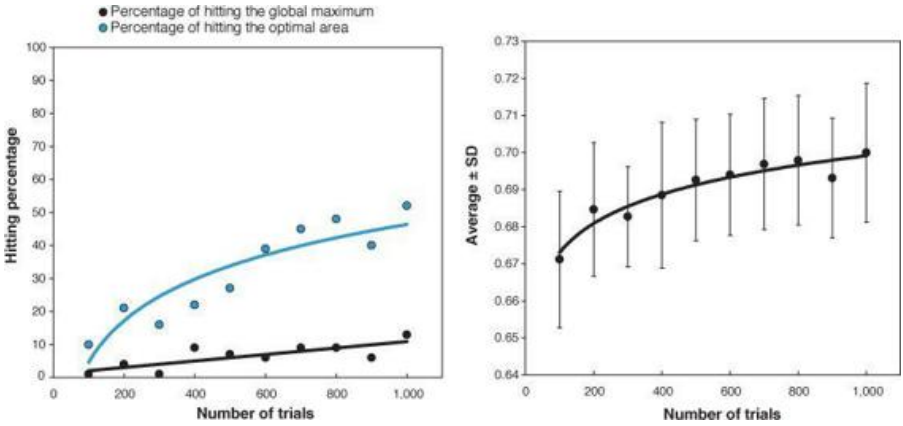

Figure 2.7.6. The relationship between the effectiveness of random search and the number of selected nodes (trials). Objective function: percentage of profitable trades. The left-hand chart shows the relationship for two characteristics: percentage of hitting the global maximum and percentage of hitting the optimal area. The right-hand chart shows the average value of the optimal solution (another effectiveness characteristic) and its variability in form of a standard error.

Random search effectiveness is higher for the optimization space corresponding to the profit-based objective function than it is for the space constructed using the function based on the percentage of profitable trades (compare the left-hand charts of Figures 2.7.5 and 2.7.6). This fully complies with the results obtained in the preceding section for direct search methods. Besides, the variability of optimal solutions estimated for the profit-based objective function decreases as the number of trials increases (see the right-hand chart of Figure 2.7.5). However this trend was not observed for the optimization space constructed using another objective function, the percentage of profitable trades (see the right-hand chart of Figure 2.7.6).

When search effectiveness is measured by the percentage of hitting the global maximum, random search is inferior to the alternating-variable ascent and the Hook-Jeeves methods. However, when a sufficiently large number of trials is executed, random search becomes more effective with two other indicators. When effectiveness is expressed through the average value of the optimal solutions or by the percentage of hitting the optimal area, random search is more effective than the alternating-variable ascent method, provided that the optimization space was constructed using the profit-based objective function and that the number of trials is equal to or greater than 400. When 500 trials are used, the random method becomes better than the Hook-Jeeves method (compare Figure 2.7.5 and Table

2.7.1 data). If optimization space was created using the objective function based on the percentage of profitable trades, random search is more effective than the alternating-variable ascent method and the Hook-Jeeves method, provided that effectiveness is measured by the percentage of hitting the optimal area and the number of trials is equal to or greater than 600. When effectiveness is measured by the average value of the optimal solution, random search is better than these two methods beginning from 700 trials (compare Figure 2.7.6 and Table 2.7.1 data).

This analysis brings about a number of important conclusions regarding the applicability of random search for optimization of trading strategies. In general, we should admit that as the number of trials increases to a certain level, the probability that the optimal solution obtained will be close enough to the global maximum becomes sufficiently high. In particular, random search may be used if (1) optimization space size allows analyzing around 20% of its nodes and (2) preliminary knowledge indicates that optimization space is unimodal. The information concerning the space modality might be available if an exhaustive search has already been implemented during the preparatory stages of strategy development. If the space shape (as determined during the preliminary investigation) is similar to the one pertaining to the profit-based objective function (see Figure 2.2.2), random search can be used in further strategy improvements and optimization.

2.8. Establishing the Optimization Framework: Challenges and Compromises

Setting up a framework for optimization of trading strategies, in general, and option strategies, in particular, is a challenging task. This requires making numerous decisions that influence both the quantity of computational resources needed to implement all necessary procedures and the reliability of the final results. The main difficulty consists in finding the trade-off between minimization of calculation time and maximization of information volume that can be obtained in the course of optimization.

In particular, the construction of the optimization space requires making several compromise decisions. The first one is about the dimensionality of optimization, which is determined by the number of parameters. Although the set of parameters requiring optimization depends on the strategy logic, the decision as to which of them are to be optimized by technical methods (similar to those described in this chapter) and which should be fixed on the basis of a priori assumptions (or scientific methods) is certainly a matter of

compromise. The fewer parameters that get optimized, the less complex the whole process is and the lower the risk is of overfitting. On the other hand, ignoring optimization (by fixing parameter values) increases the risk to miss the profitable variant of the strategy. The next decision consists of choosing the permissible parameter ranges and optimization steps. Here, the trade-off is only between the time consumed and the volume of information obtained. Therefore, these decisions depend mostly on computational facilities available to the strategy developer.

Reduction of calculation time is essential but it is not the only challenge arising for the strategy developer. Another fundamentally important decision is the composition, quantity, and relative importance of objective functions utilized in the optimization process. We recommend basing this decision on the extent of correlations between different objective functions and on the scope of additional information contained in them. After the set of objective functions is defined, we need to choose a multicriteria analysis method, which is not a trivial problem. The choice of particular direct search method is also one of the main trade-offs (solved on the basis of analysis of optimization space properties) that influence the reliability of optimization results.

In this chapter we used the example of the basic delta-neutral strategy to demonstrate general approaches to resolving trade-offs and making key decisions in the process of optimization. While discussing the main elements of optimization framework of the delta-neutral strategy, we strove to put them in the general context of option strategies optimization. Since each optimization procedure depends on specific properties of a particular strategy, we could not provide universal solutions that will be equally suitable for all types of options strategies. Instead, we tried to develop a system of recommendations that will enable the developer to construct a reliable system suitable for optimization of *different* option strategies.

Chapter 3. Risk Management

As of today, there is no universal definition that would embrace all aspects of such a complex and versatile concept as “risk.” Actually, we must admit that millions of people—journalists, businessmen, scientists, professional investors, and financial services consumers who use this term daily—are still unable to give a strict and universal definition of the matter in question. This is especially striking considering that the notion of risk is the cornerstone of economics, finance, and many other related disciplines. Moreover, the same situation is observed in two other, undoubtedly fundamental, areas of scientific cognition of the material world: biology and physics. In biology, there is no universal definition of “species,” while this notion is the basis for the whole macroevolution theory—the signal achievement of this science. Physicists also did not arrive at a common opinion regarding a unique and comprehensive definition of “energy.” Without precise understanding of this key element, neither the development of the quantum theory at the microcosmic level nor the creation of the “final theory of everything,” pretending to describe the origin, evolution, and future fate of the universe, is possible.

Isn’t it strange that the three basic areas of human knowledge—biology, economics, and physics—erect their theories on the basis of core elements lacking strict and unambiguous scientific definitions? We leave this question unanswered since even the slightest attempt to straighten out this confusing situation will distract us not only from the main topic, but also from the system of rational reasoning that we rigorously adhere to in this book.

3.1. Payoff Function and Specifics of Risk Evaluation

All financial instruments can be classified as linear or nonlinear according to their payoff function. The former category includes stocks, commodities, currencies, and other assets whose profits and losses are directly proportional to their price. Nonlinear assets include derivative financial instruments with prices depending on the prices of their underlying assets. The relationships between the profits and losses of these instruments and the underlying asset prices are not linear. Options represent one of the main categories of the nonlinear instruments. The approaches used to evaluate the risks of linear and nonlinear instruments differ fundamentally.

3.1.1. Linear Financial Instruments

The grounds for risk management theory were established when derivatives had not yet begun their full-scale advance on the financial markets. As a result, all classical risk evaluation models were developed for linear instruments. The general concept stating that "the risk of a specific asset is proportional to the variability of its price" was accepted as the basic paradigm for risk quantification.

The objective estimation of price variability can be derived only from the past price fluctuations (other estimates, for example, those based on expert opinions, cannot be completely objective). Such an approach has a significant drawback, since it requires extrapolation of historical data and is based on the assumption that probabilities of future outcomes are proportional to the frequency of similar occurrences that have been realized in the past. Although in many areas (for example, in car insurance) this method is widely accepted, it was repeatedly proved that with regard to financial markets it is, to say the least, imperfect. Nevertheless, despite all the drawbacks and inaccuracies that arise when risks are estimated on the basis of historical data, this approach is widely applied because there are no better alternatives as of today.

It was suggested to use the standard deviation of asset returns (or their dispersion) as the measure of price variability, which is usually denoted by the term "historical volatility." Usually, historical volatility is calculated as annualized mean-square deviation of daily logarithm price returns. The length of the historical period used to calculate volatility is a key parameter in measuring the risk of linear assets. If time series are too long, the estimate of the current risk is based on outdated data having no direct relation to the actual market dynamics. Such a valuation cannot be reliable. On the other hand, using time series that are too short brings about unstable risk estimates reflecting only momentary market trends. Thus, the choice of the historical horizon length is a product of compromise and should be defined within the context of the whole risk management system (while taking into account the specific features of the trading strategy under development).

Although standard deviation by itself reflects the risk, it is also used to calculate more complex indicators that express risks in a form that is more convenient for practical application. The well-known example of such an indicator is Value at Risk (VaR), which estimates an amount of loss that will not be exceeded with a given probability during a certain period.

The history of emergence and broad acceptance of this indicator date back to the stock market crash of 1987, when the failure of existing risk management mechanisms became evident. The search for

new approaches to risk measurement led to quick development and an extensive application of the innovative technology, expressing the risk through estimation of capital that may be lost with a given probability, rather than by calculation of some statistical indicators (like standard deviation). In the 1990s VaR became the universally recognized standard of risk measurement, and in 1999 it obtained the official international status set in the Basel Agreements. With the lapse of time, VaR became the compulsory indicator appearing in the accounting reports of the majority of financial institutions.

Although from a practical point of view, VaR is more informative than standard deviation, these indicators do not differ from each other conceptually. This can be demonstrated by considering the VaR calculation method. There are three main approaches to VaR calculation: analytical, historical, and the Monte-Carlo method. *The analytical method* is based on using the parameters of specific return distribution. Despite its numerous drawbacks, lognormal distribution is used most often. Since the main parameter of this distribution is the standard deviation (the second parameter, expected price, is usually set to be equal to the current asset price), one can claim that VaR is just a secondary indicator derived from the standard deviation. *The historical method* is based on using asset price changes that were observed during a predetermined period in the past. Since the standard deviation is calculated on the basis of the same data, both indicators strongly correlate with each other and, in fact, express the same concept. *The Monte-Carlo method* generates a multitude of random price (or return) outcomes. The algorithm of price generation utilizes the probability density function of a certain distribution. And again, the lognormal distribution (with the standard deviation as its main parameter) is used most frequently. Thus, the emergence of VaR added more convenience for its users, but did not lead to the development of new risk evaluation principles.

There is nothing particularly complicated in risk evaluation for portfolios consisting of linear assets. In most cases the standard deviation and VaR of such portfolios can be calculated using simple analytical methods. The minimal input data required for these calculations include the standard deviation of each instrument, its weight in the portfolio, and the covariance matrix. The matrix is necessary to account for the effect of diversification (a decrease in portfolio risk as a result of including low-correlated assets or assets with negative correlations in it). Conversely, the risk for a portfolio containing at least some nonlinear instruments cannot be calculated analytically. In this case we need to apply different simulation methods, of which the Monte-Carlo is the most common.

3.1.2. Options as Nonlinear Financial Instruments

Traditional risk evaluation methods that are commonly used for linear assets are inappropriate for financial instruments with a nonlinear payoff function. The reason is that the return distribution of nonlinear assets is not normal. For example, a call option has an unlimited profit potential which implies that the right tail of its return distribution is unlimited. At the same time, its maximum possible loss cannot exceed the premium paid at the position opening. Hence, the left tail of the distribution is limited to this amount, which means that the returns distribution is asymmetric and cannot even roughly be considered as normal. The nonnormality is so pronounced that the application of logarithmic transformation does not solve the problem, as this is the case for linear assets. (Although logarithmic transformation brings the return distribution of linear assets closer to normal distribution, there are many evidences of deviation of logarithm return distribution from normality. Nevertheless, in this case we speak only about deviations, while for nonlinear assets the distribution does not even get close to the normality.)

Despite the aforesaid, methods developed for risk measurement of linear assets can be applied to some nonlinear instruments, provided that certain conditions are observed. For example, although the VaR of an option cannot be calculated analytically, it can be estimated using the Monte-Carlo method. However, these methods can be used only as auxiliary instruments for the risk evaluation of automated option trading strategies. Dedicated methods with due regard to the specifics of nonlinear assets in general and options in particular should be given priority.

The common tools for risk evaluation of separate options are the “Greeks” that express the change in option price given the small change in a specific variable. These indicators can be interpreted as the sensitivity of an option to fluctuations in a variable value. The variables include an underlying asset price, volatility, time, and the interest rate. These variables may be seen as risk factors that give rise to instability of option prices.

The Greeks are calculated analytically as partial derivatives of the option price with respect to the given variable. One of the option pricing models (for example, the Black-Scholes formula) is used to calculate these derivatives. Delta is the derivative of the option price with respect to the underlying asset price. The derivative with respect to volatility is called vega, with respect to time—theta—and with respect to the interest rate—rho. Derivatives of the second and higher orders can also be used (for example, gamma—the second order derivative with respect to price).

The Greeks represent a convenient and rather adequate instrument of options risk evaluation. The risk of combination, consisting of options related to one underlying asset, can be estimated by summing up the corresponding Greeks. However, *problems do arise in measuring the risks of complex structures rather than separate options.* Inclusion of option combinations, related to different underlying assets, in the portfolio makes it impossible to evaluate certain risks by mere summation. Some Greeks, such as theta and rho, are additive for options pertaining to different underlying assets. Thus, the sensitivity of the complex portfolio to time decay or the interest rate change is easily calculated as the sum of thetas and rhos of separate options. Things get more complicated with non-additive risk indicators, such as delta and vega (these two characteristics are the most important for risk management). If the portfolio consists of options relating to several underlying assets, the summation of separate deltas and vegas makes no sense. These characteristics are not appropriate for complex portfolios since they express changes in the option price depending on the changes in price or volatility of one specific underlying asset. Differentiated analysis of positions in separate underlying assets is not efficient since *the whole portfolio must be evaluated as a single structure.*

The problem of Greeks non-additivity can be solved by expressing the deltas of all options as the derivatives with respect to some common index rather than to their corresponding underlying assets. Similarly, vegas of different options can be calculated as the derivative with respect to the volatility of the same index rather than to the volatilities of separate underlying assets. Such procedures impart additivity properties to delta and vega, which allows for the calculation of risk indicators for the whole portfolio as a single entity. S&P 500 or any other index constructed by the system developer (for example, an index based on the prices of only those stocks that are underlying assets for combinations included in the portfolio) can be used as a common index. The selection or construction of the common index represents a separate complex task; its solution depends on the relationships between the portfolio elements and the chosen index and on many risk management parameters.

Index delta is undoubtedly one of the most important instruments for measuring the risk of option portfolios. However, apart from this indicator, risks of automated options trading strategies should and must be evaluated using additional complementary characteristics (in this chapter, apart from index delta, we discuss three other indicators: VaR, loss probability, and asymmetry coefficient). This enables us to put together a complete picture of different risk aspects that are peculiar to the strategy under development.

3.2. Risk Indicators

The risks of option portfolios may be estimated using traditional indicators, such as VaR, and specific characteristics developed deliberately for this purpose. Although the risk evaluation of option portfolios by means of indicators that are usually used for linear assets is technically possible, one should be extremely careful in interpreting the obtained results. These indicators can be used only as auxiliary instruments of risk management. The main indicators should be developed subject to options peculiarities. To get a thorough comprehension of different risks of the portfolio under construction, several indicators should be applied simultaneously. Moreover, these indicators should correlate with each other as little as possible. This will allow us to form a set of unique risk indicators, each of which will complement, but not duplicate, the information contained in other indicators.

Four risk indicators are discussed in this section. We begin with VaR and explain how this indicator can be calculated for the option portfolio. Then, we describe three indicators developed specifically for risk evaluation of complex option portfolios. Particular attention is focused on the index delta.

3.2.1. Value at Risk (VaR)

As defined earlier, VaR is an estimate of a loss that will not be exceeded with a given probability over a certain period (in further examples we will use the 95% probability). For a portfolio consisting of nonlinear assets, the Monte-Carlo simulation has to be applied for a VaR calculation. A multitude of underlying asset price outcomes is generated using the probability density function of lognormal (or another) distribution (constructed on the basis of standard deviation derived from historical data). Then, for each price outcome the option payoff is estimated. Finally, these values are used to calculate VaR. The same method (with an allowance for correlations between different underlying assets) is used to calculate the VaR of the option portfolio.

Despite all its advantages and practical utility, VaR has a number of significant drawbacks (Tsudikman, Izraylevich, Balishyan, 2011), which become especially apparent when it is applied to assess the risks of nonlinear assets. The following are the main drawbacks:

- Lack of reliable technology to derive robust parameters of return distribution from historical data. Consequently, the probability density functions used for VaR estimation fail to forecast severe economic shocks. For complex portfolios that

include options relating to different underlying assets, the possibility of obtaining reliable and robust VaR estimates is even more questionable.

• Besides underestimating the magnitude and the frequency of extreme outcomes, VaR may also overestimate both unsystematic (diversifiable) and systematic risks.

• VaR is a point estimate and thus does not reflect the whole range of potential outcomes. Since it is calculated for only one or several probability values, the general properties of the distribution left tail (containing all adverse outcomes) remain vague.

• An estimated VaR value can be easily manipulated by changing the length of the historical price data. This issue becomes especially problematic if the portfolio includes less liquid securities (like many options) lacking a verified price history.

• Like many other indicators developed for estimating the risk of linear assets, VaR relies, to a certain extent, on effective market hypothesis and, hence, inherits its numerous drawbacks. This becomes especially apparent in times of extreme market fluctuations.

Until recently it was widely believed that VaR adequately describes the risk of all investment portfolios regardless of their composition and structure. However, the global financial crisis that erupted in 2007 vividly demonstrated the mismatch of forecasts based on VaR and realized losses. The reason for the divergence between forecasts and reality was that in the past 20 years financial markets underwent cardinal changes, caused by the development of complex financial technologies and the shifting of emphasis from simple linear assets to derivatives (many of which are nonlinear). The share of nonlinear instruments (of which options are not the last) in the asset structure of large financial institutions grew steadily. This impressive advance notwithstanding, risk evaluation mechanisms remained the same or lagged behind significantly.

In the development of option-based trading strategies, the use of a risk forecasting system based solely on VaR or similar dispersion-related indicators is inadmissible. A fully fledged risk management system should include the whole package of evaluation algorithms based on various principles and take specific option properties into account.

3.2.2. Index Delta

The index delta characterizes the sensitivity of an option portfolio to broad market fluctuations. This indicator expresses quantitatively the change in the portfolio value under a small index change. The index delta can be used to evaluate and manage the risk of complex portfolios in the same way as an ordinary delta does (when applied for portfolios consisting of options related to a single underlying asset). Besides, the index delta can be applied to create delta-neutral portfolios (as described in Chapter 1, "Development of Trading Strategies") and to restructure existing positions in order to maintain delta-neutrality over the whole period of portfolio existence.

The Algorithm

The algorithm for calculation of the index delta can be presented as a sequence of the following steps:

1. Build regression models for the relationships between the prices of all underlying assets (options on which are included in the portfolio) and the index. To build these models, the length of the historical horizon has to be specified.

2. Using the regression models created at step 1, calculate the prices of all underlying assets as implied by the one-point change in the index value (or by a certain percentage index change).

3. Determine the fair value of each option in the portfolio given that its underlying asset price is equal to the value obtained at step 2. This procedure is executed by substituting the price obtained at step 2 (rather than the current price) into the selected option pricing model.

4. For each option in the portfolio, calculate the price increment that would take place if its underlying asset price changes by one point (or by the other value defined at step 2). The price increment is the difference between the value calculated at step 3 and the current option market price.

5. Calculate the value of the index delta by summing up all the increments obtained at step 4.

Analytical Method of Calculating Index Delta

Consider a portfolio consisting of options related to different underlying assets. Let $\{O_1, O_2, O_3,\ldots,O_N\}$ be a set of options included in the portfolio, while the underlying asset of option O_i is A_i. If different options relate to the same underlying asset (that is, assets A_j and A_k coincide under $j \neq k$), these options form a structure that may correspond to one of the standard option combinations (strangle,

straddle, calendar spread, and so on) or may be of any arbitrary composition. Formally, the combination may consist of an unlimited number of different options relating to one underlying asset, and the portfolio may include an unlimited number of combinations.

For any option O_i included in the portfolio, its delta can be expressed as

$$\Delta_i = \left(\frac{\partial O_i}{\partial A_i} \right),$$

where ∂O_i and ∂A_i denote small changes in the price of the option and its underlying asset, respectively. This expression interprets delta as the speed of the option price change relative to price changes of its underlying asset.

Computing delta Δ_i of a separate option is a rather trivial task. It is realized in many software programs based on the Black-Scholes and other, more sophisticated, models. The delta of the portfolio consisting of options on different underlying assets cannot be computed as the sum of all deltas since they are derivatives of premiums with respect to different variables (prices of different underlying assets). As noted previously, this problem is solved by calculating the sensitivity of the option price relative to the changes of index rather than to changes in the prices of separate underlying assets. Following this approach, we define index delta as a derivative of the option price with respect to the index value:

$$ID_i = \left(\frac{\partial O_i}{\partial I} \right).$$

This index delta of a separate option can be expressed as follows:

$$ID_i = \left(\frac{\partial O_i}{\partial I} \right) = \left(\frac{\partial O_i}{\partial A_i} \right) \cdot \left(\frac{\partial A_i}{\partial I} \right). \qquad (3.2.1)$$

In this equation the value expressed as $\left(\frac{\partial A_i}{\partial I} \right)$ represents the change in the price of the underlying asset in response to the index change. In this respect it is similar to the concept of beta. The difference between these two characteristics is that the index delta is dimensional, whereas beta is a nondimensional characteristic (expressed as the ratio of relative price changes to a relative index change). Beta is commonly used to evaluate the relationship between the changes in the index and a separate asset. For our purposes, it

would be convenient to express it as the following ratio:

$$\beta_i = \frac{\partial A_i / A_i}{\partial I / I}.$$

After simple transformations we obtain

$$\left(\frac{\partial A_i}{\partial I}\right) = \left(\frac{\partial A_i / A_i}{\partial I / I}\right) \cdot \frac{A_i}{I} = \beta_i \cdot \frac{A_i}{I}.$$

By substituting this expression into equation 3.2.1, we get the formula for calculating the index delta of a separate option,

$$ID_i = \Delta_i \cdot \beta_i \cdot \frac{A_i}{I}, \qquad (3.2.2)$$

where Δ_i is the ordinary delta of option O_i. Let us denote the quantity of option O_i in the portfolio as x_i. To calculate the index delta of the whole portfolio $ID_{Portfolio}$, the index deltas of separate options included in this portfolio should be summed considering their quantities:

$$ID_{Portfolio} = \sum_i \Delta_i \cdot \beta_i \cdot A_i \cdot x_i / I. \qquad (3.2.3)$$

The index delta calculated using equation 3.2.3 can be interpreted as the change in the portfolio value in response to the index change by one point. It might be more convenient to express the portfolio value change in response to the percentage index change. This allows evaluating portfolio sensitivity to relative index changes. For example, a simple transformation of equation 3.2.3 allows calculating the index delta for a 1% index change:

$$ID^{\%}_{Portfolio} = \frac{1}{100} \sum_i \Delta_i \cdot \beta_i \cdot A_i \cdot x_i. \qquad (3.2.4)$$

Example of Index Delta Calculation

Let us consider an example of calculating the index delta for a small portfolio, consisting of options on seven stocks. The S&P 500 will be used in the calculations, though other indexes could be equally appropriate here. Table 3.2.1 shows the portfolio consisting of seven short straddles. The quantities of these combinations are approximately inversely proportional to the prices of the corresponding underlying stocks. The portfolio was created on January 2, 2009, using options with the nearest expiration date (January 16, 2009). The current index value on January 2, 2009, was 931.8. Beta coefficients of the stocks were calculated on the basis of daily closing prices using the 120-day historical horizon. Ordinary deltas of separate options were derived using the Black-Scholes model

with a risk-free interest rate of 3.3% and the volatility estimated at the same horizon of price history.

Table 3.2.1. Data required to calculate the index delta (relative to the S&P 500 index) of the portfolio consisting of seven short straddles.

Stock Ticker	Stock Price	Stock Beta	Strike	Option Type	Quantity	Delta of One Option	Index Delta of One Option	Index Delta of the Position
VLO	23.24	1.58	22.5	Call	−400	0.63	0.0248	−9.93
				Put	−400	−0.37	−0.0146	5.83
PPG	43.55	1.01	45	Call	−200	0.39	0.0184	−3.68
				Put	−200	−0.60	−0.0283	5.66
VNO	58.29	1.62	60	Call	−100	0.46	0.0466	−4.66
				Put	−100	−0.55	−0.0557	5.57
SYMC	14.80	0.91	15	Call	−600	0.46	0.0066	−3.99
				Put	−600	−0.53	−0.0077	4.60
CBE	30.67	1.02	30	Call	−300	0.61	0.0205	−6.14
				Put	−300	−0.40	−0.0134	4.03
NKE	53.06	0.89	55	Call	−100	0.35	0.0177	−1.77
				Put	−100	−0.64	−0.0324	3.24
MMM	59.19	0.70	60	Call	−100	0.44	0.0196	−1.96
				Put	−100	−0.55	−0.0245	2.45
Total (index delta of the portfolio)								−0.61

The next-to-last column of Table 3.2.1 shows index deltas of separate options calculated by equation 3.2.2. For example, the index delta of one call option related to VLO stock is calculated as $IDi = (23.24 \times 1.58 \times 0.63)/931.8 = 0.0248$. The product of index delta of one option and its quantity in the portfolio gives the index delta of the position corresponding to this contract (the last column of the table). Thus, the index delta of the position corresponding to the call of VLO stock is $IDi = -400 \times 0.0248 = -9.93$. Summing the index deltas of all positions (that is, all values shown in the last column of the table) gives the index delta of the portfolio, which corresponds to equation 3.2.3. In this example $IDPortfolio = -0.61$. Using equation 3.2.4, the portfolio index delta can easily be expressed in percentage terms: $ID^{\%}_{Portfolio} = (-0.61 \times 931.8)/100 = -5.69$, which means that the portfolio value will decrease by 5.69% if the S&P 500 changes by 1%.

Analysis of Index Delta Effectiveness in Risk Evaluation

To estimate the effectiveness of the index delta, we used a database containing eight years of the price history of options and their underlying assets. The index deltas were estimated relative to the S&P

500 index. Accordingly, stocks composing this index were used as underlying assets for options included in the portfolio.

To estimate the influence of portfolio creation timing on the index delta effectiveness, we constructed a series of portfolios for each expiration date from the beginning of 2001 until the beginning of 2009. These portfolios differed from each other in the number of trading days between the time of portfolio creation and the expiration date. For a given expiration date, the most distant of the portfolios was constructed 60 days before the expiration; the next one, 59 days before the expiration; and so on, right up to the last portfolio that was created only 1 day before the expiration. Thus, 59 portfolios (different from each other in terms of time left until options expiration) were constructed for each expiration date. In total, from 30 (for 60 days) to 90 (for 1 day) portfolios were created for each value of the "number of days left until options expiration" parameter.

Each portfolio consisted of short straddles for all 500 stocks included in the index. The straddles were created using the strikes closest to the current stock price. The quantity of options corresponding to each stock was determined as $x_i = 10000 / A_i$, where 10000 represents the equivalent of the capital allocated to each straddle (see Chapter 4, "Capital Allocation and Portfolio Construction," for details), and A is the price of the stock i. The beta of each stock, deltas of separate options, and index deltas were calculated using the same techniques and parameters as in the preceding example (see Table 3.2.1). In addition, the following characteristics were estimated for each portfolio:

- *Percent change of the index from the time of portfolio creation until the expiration date:*

  $I_{\%} = 100 \cdot (I_e - I_t)/I_t$,

  where I_t is the index value at the time t of portfolio creation, and I_e is the index value at the expiration date.
- *Percent change of the portfolio value from the time of its creation until the expiration date:*

  $$P^{\%}_{realized} = 100 \cdot (P_e - P_t)/P_t,$$

  where P_t is the market value of the portfolio at the time t, and P_e is the value of the portfolio at the expiration date. This characteristic reflects the realized portfolio risk.
- *Expected percent change of the portfolio value:*

$$P^{\%}_{expected} = 100 \cdot (I^{\%} \cdot ID^{\%}_{Portfolio}) \Big/ P_t .$$

This characteristic represents the estimate of portfolio risk given the index change.

- *The difference between the actual and the expected change of portfolio value:*

$$Difference = P^{\%}_{realized} - P^{\%}_{expected} .$$

The closer this characteristic is to zero (that is, the lesser the difference between the actual and the expected changes), the more accurate the index delta is in forecasting portfolio value fluctuations.

Figure 3.2.1 shows the average differences between the realized changes in portfolio values and the changes that were expected given the portfolio risk estimated by the index delta. Differences of the portfolios created long before the expiration (30 to 60 days) were close to zero, though their variability (shown on the chart as standard deviations) was substantial. Therefore, these portfolios allowed for the most accurate estimation of the risk with relatively high dispersion of individual outcomes. For portfolios created shortly before the expiration (up to 20 days), the difference between the actual and the expected value changes were high, positive, and rather stable (low dispersion). This implies that the index delta underestimates the risk of such portfolios and the magnitude of this underestimation is stable.

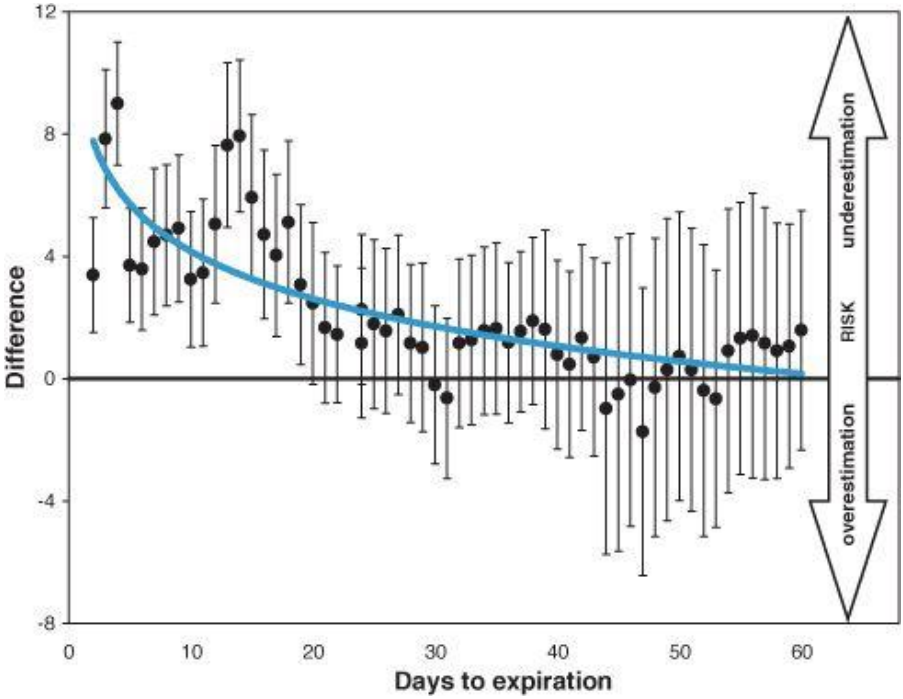

Figure 3.2.1. Relationship between the effectiveness of the index delta (expressed through the difference of the actual and the expected portfolio value changes) and the number of days left until options expiration at the moment of portfolio creation. Points denote average values; horizontal bars, standard deviations.

Positive differences $\left(P^{\%}_{realized} > P^{\%}_{expected}\right)$ imply that actual changes in the portfolio value were higher than expected. This means that the index delta underestimated the risk of these portfolios (since the loss of short positions is incurred when options values increase $(P^{\%}_{realized} > 0)$. Correspondingly, negative differences $\left(P^{\%}_{realized} < P^{\%}_{expected}\right)$ imply that actual changes in the portfolio value were smaller than expected (risk was overestimated). Bars depicting the variability of outcomes are situated in both the negative and the positive areas for portfolios created far from their expiration dates (see Figure 3.2.1). This indicates that some of these portfolios were overestimated and some were underestimated.

Overall, we can conclude that at the moment of portfolio creation, the more days that remain until the expiration date, the more accurate the average risk estimate is. At the same time, the probability that a

given estimate will turn out to be inaccurate increases as well. For portfolios created shortly before the expiration date, the risk is underestimated, though the magnitude of this underestimation is relatively stable (the standard deviation of the average difference is rather low). Therefore, for portfolios created a long time before the expiration, the effectiveness of the index delta can be increased by applying additional risk indicators, while for portfolios created just before the expiration date, the introduction of adjusting coefficients seems to be sufficient (since the variability of results is low).

Delta is a *local* instrument indicating the change in the option price under a *small* change in its underlying asset price. This means that the higher the actual underlying price change, the less accurate the forecast of the option price change based on the ordinary delta is. Hence, the next issue addressed in our study is to test whether and by how much the effectiveness of the forecast based on the index delta deteriorates when index changes are considerable. To answer this question, we will examine the relationship between the effectiveness of the index delta (expressed through differences between actual and expected changes in the portfolio value) and the magnitude of the index change.

As it follows from Figure 3.2.2, big index changes indeed induce considerable differences between realized and expected changes in portfolio values. The relationships are statistically significant in both cases when the index grows and when it falls (in the former case the correlation was lower than in the latter case). The most interesting fact in the context of evaluation of the index delta effectiveness is that considerable index changes (regardless of whether it increases or decreases) correspond to positive differences, and small index changes correspond to negative ones (see Figure 3.2.2). This means that *under significant index fluctuations, the index delta tends to underestimate the risk, whereas under small and moderate index moves, the risk turns out to be overestimated.* The index delta demonstrates the highest effectiveness when the amplitude of market fluctuations is within the limits of 3% to 5%.

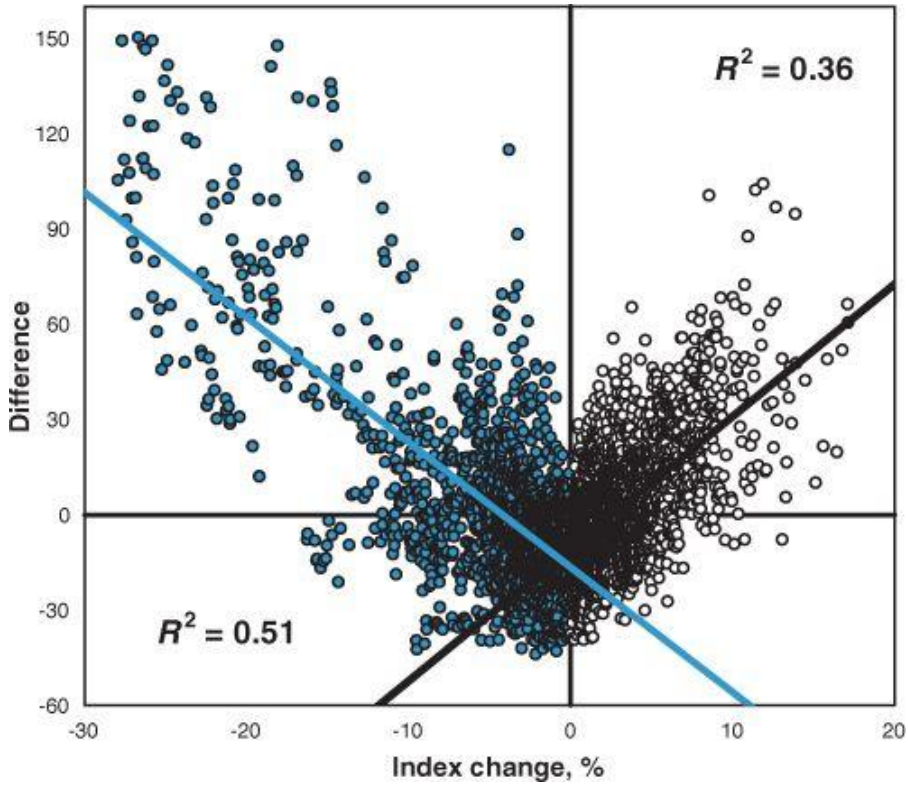

Figure 3.2.2. Relationship between the effectiveness of the index delta (expressed through the difference of the actual and the expected portfolio value changes) and the percentage index change. Empty points correspond to positive index changes; filled points, to negative index changes.

Analysis of Index Delta Effectiveness at Different Time Horizons

The research presented in the preceding section was limited to the cases when risk is estimated only once—at the time of portfolio creation. The accuracy of this estimate was also verified only once—at the expiration date of options included in the portfolio (for simplicity's sake we assume that all options expire simultaneously). In this section we will study the situations in which risk is estimated repeatedly during the portfolio life span and the effectiveness of this estimate is tested at different time intervals.

Contrary to the previous study (in which 59 portfolios were formed for each expiration date), in this study a single portfolio was constructed for each expiration date. The time of each portfolio creation was 50 trading days away from the given expiration date. The number of portfolios totaled 90. All other parameters of the trading strategy and the algorithm used for portfolio creation were the

same as in the previous study.

The effectiveness of the index delta in forecasting portfolio risk was evaluated using the method that was applied in the preceding section (the difference between risk estimated on the basis of the index delta and its realized value). To assess the quality of risk forecast at different stages of portfolio existence, we had to (1) calculate values of the index delta daily through the whole life span of the portfolio, and (2) estimate the changes in the portfolio value during different time periods (we will refer to these periods as "testing horizons"). All periods, from the shortest (1 day) to the longest ones (49 days) had to be tested. This data allows us to perform a detailed evaluation of the differences between forecasts and reality.

Table 3.2.2 presents an example of one of the portfolios and shows the evolution of its characteristics. At the time of portfolio creation (August 7, 2008; first row of the table), there were 50 trading days until the options expiration date (October 17, 2008). The second row of the table shows characteristics of this portfolio the day after its creation. Accordingly, the third row shows its characteristics after three days, and so on, until the expiration date (only 20 first days of portfolio existence are included in Table 3.2.2). The table shows the values of the following characteristics that are necessary to estimate the effectiveness of the index delta at different time horizons:

- d—The number of days from the valuation moment until the expiration date. This characteristic is used for indexing the moments of valuation and testing.
- ID_d—The index delta calculated at the moment of valuation using equation 3.2.3.
- $ID_d^{\%}$—The percentage of index delta (expresses the change in the portfolio value if the index changes by 1%) calculated at the moment of valuation using equation 3.2.4.
- $I_{\%}$—The percentage of index change calculated as $I_{\%} = 100 \cdot (I_{d-j} - I_d)/I_d$, where I_d is the index value at the moment of valuation, I_{d-j} is the index value at the moment of testing, and j is the testing horizon (number of days between the valuation date and the testing date).
- $P_{realized}^{\%}$—The percentage of change of the portfolio value calculated as $P_{realized}^{\%} = 100 \cdot (P_{d-j} - P_d)/P_d$, where P_d is the market value of the portfolio at the moment of valuation d, and P_{d-j} is the value of the portfolio at the moment of testing. This characteristic reflects the change in

the portfolio value that occurs during j days. Therefore, $P^{\%}_{realized}$ represents the risk of the portfolio that was realized at the specified time horizon.

- $P^{\%}_{expected}$—The expected percentage change of portfolio value calculated as $P^{\%}_{expected} = 100 \cdot (I^{\%} \cdot ID^{\%}_{d}) / P_{d}$. This characteristic estimates the portfolio risk under the condition that during j days the index changes by $I\%$.

- *Difference*—A reflection of the divergence between the actual and the expected change in the portfolio value $\left(P^{\%}_{realized} - P^{\%}_{expected}\right)$. A positive difference means that the index delta underestimates the risk, whereas a negative difference points to the risk overestimation. The closer the difference is to zero, the more accurate the index delta is in forecasting the portfolio risk.

Table 3.2.2. Characteristics of the option portfolio measured during 20 days from the time of its creation (August 7, 2008). Values presented in the table are used to evaluate the effectiveness of the index delta at different time horizons.

Date	d	P_d	D_d	$D^{\%}_d$	I_d	Testing Horizon j = 1 Day				Testing Horizon j = 5 Days			
						$I^{\%}$	$P^{\%}_{realized}$	$P^{\%}_{expected}$	Difference	$I^{\%}$	$P^{\%}_{realized}$	$P^{\%}_{expected}$	Difference
7 Aug 08	50	1,224	−0.59	−7.5	1,266								
8 Aug 08	49	1,237	−1.38	−17.8	1,296	2.4	1.1	−1.5	2.5				
11 Aug 08	48	1,267	−1.74	−22.8	1,305	0.7	2.4	−1.0	3.4				
12 Aug 08	47	1,243	−1.32	−17.0	1,290	−1.2	−1.9	2.2	−4.1				
13 Aug 08	46	1,225	−1.16	−14.9	1,286	−0.3	−1.4	0.4	−1.8				
14 Aug 08	45	1,235	−1.48	−19.1	1,293	0.6	0.8	−0.7	1.4	2.1	0.9	−1.3	2.2
15 Aug 08	44	1,235	−1.68	−21.8	1,298	0.4	0.0	−0.6	0.6	0.1	−0.2	−0.2	0.0
18 Aug 08	43	1,183	−1.13	−14.5	1,279	−1.5	−4.2	2.7	−6.9	−2.0	−6.7	3.7	−10.3
19 Aug 08	42	1,159	−0.73	−9.2	1,267	−0.9	−2.0	1.1	−3.2	−1.8	−6.7	2.4	−9.2
20 Aug 08	41	1,151	−0.82	−10.5	1,275	0.6	−0.7	−0.5	−0.2	−0.9	−6.1	1.1	−7.2
21 Aug 08	40	1,148	−0.87	−11.1	1,278	0.2	−0.3	−0.2	−0.1	−1.2	−7.1	1.8	−8.9
22 Aug 08	39	1,126	−1.27	−16.5	1,292	1.1	−1.9	−1.1	−0.8	−0.5	−8.8	0.8	−9.6
25 Aug 08	38	1,100	−0.58	−7.3	1,267	−2.0	−2.3	2.9	−5.2	−0.9	−7.0	1.1	−8.1
26 Aug 08	37	1,087	−0.66	−8.4	1,272	0.4	−1.3	−0.2	−1.0	0.4	−6.3	−0.3	−6.0
27 Aug 08	36	1,074	−0.98	−12.5	1,282	0.8	−1.1	−0.6	−0.5	0.6	−6.7	−0.5	−6.2
28 Aug 08	35	1,071	−1.51	−19.7	1,301	1.5	−0.3	−1.7	1.4	1.8	−6.7	−1.7	−5.0
29 Aug 08	34	1,060	−1.08	−13.8	1,283	−1.4	−1.0	2.5	−3.5	−0.7	−5.8	1.1	−6.9
2 Sep 08	33	1,062	−0.98	−12.5	1,278	−0.4	0.1	0.5	−0.4	0.8	−3.5	−0.6	−3.0
3 Sep 08	32	1,078	−0.96	−12.2	1,275	−0.2	1.6	0.2	1.3	0.3	−0.8	−0.2	−0.6
4 Sep 08	31	1,088	0.08	1.0	1,237	−3.0	0.9	3.4	−2.5	−3.5	1.2	4.1	−2.9

The characteristics shown in Table 3.2.2 correspond to two testing horizons, one and five days. Let us consider as an example the 41st

day to options expiration ($d = 41$, highlighted in gray in the table). The following is a step-by-step description of the calculations necessary to estimate the difference between realized and expected risks. The index delta corresponding to this date is equal to

$$ID^{\%}_{41} = (I_{41} \cdot ID_{41})/100 = (1275 \cdot -0.82)/100 = -10.5 .$$

For the one-day testing horizon ($j = 1$) the percentage index change is estimated as

$$I\% = 100 \cdot (I40 - I41)/I41 = 100 \cdot (1278 - 1275)/1275 = 0.2\%.$$

The percentage change of the portfolio value is calculated as

$$P^{\%}_{realized} = 100 \cdot (P_{40} - P_{41})/P_{41} = 100 \cdot (1148 - 1151)/1151 = -0.3\% ,$$

and the expected percentage change of the portfolio value is

$$P^{\%}_{expected} = 100 \cdot (I^{\%} \cdot ID^{\%}_{Portfolio})/P_{41} = 100 \cdot (0.2 \cdot -10.5)/1151 = -0.2\%.$$

The difference between the actual and the expected changes in the portfolio value is equal to *Difference* = –0.3 – (–0.2) = –0.1%, which indicates that on the 41st day the risk was slightly overestimated (when the estimate is tested just one day after it has been made). For a five-day testing horizon ($j = 5$), the percentage index change is

$$I\% = 100 \cdot (I36 - I41)/I41 = 100 \cdot (1282 - 1275)/1275 = 0.6\%.$$

The percentage change of the portfolio value is equal to

$$P^{\%}_{realized} = 100 \cdot (P_{36} - P_{41})/P_{41} = 100 \cdot (1074 - 1151)/1151 = -6.7\% ,$$

and the expected portfolio value change is

$$P^{\%}_{expected} = 100 \cdot (I^{\%} \cdot ID^{\%}_{Portfolio})/P_{41} = 100 \cdot (0.6 - 10.5)/1151 = -0.5\%.$$

The difference between the actual and the expected changes in the portfolio value is much higher: *Difference* = –6.7 – (–0.5) = –6.2% , which implies that when the risk is tested five days after the forecast, it becomes highly overestimated.

Characteristics of all portfolios created from the beginning of 2001 until the beginning of 2009 were calculated using the same methodology. The index delta of each portfolio was estimated daily during the whole period of portfolio existence. The effectiveness of these estimates was tested at all time horizons (from 1 to 49 days).

Graphical presentation of the relationship between the realized and the forecast changes in the portfolio value provides a visual insight into the effectiveness of the index delta. Figure 3.2.3 shows such relationships for the initial stage of the portfolio life (from the 50th to the 40th days until options expiration). The highest correlation was detected for the one-day testing horizon. In this case the cloud of

points (each of which represents an individual portfolio) is blurred to the smallest extent (the highest determination coefficient $R2 = 0.29$ was obtained in this case). The extension of the testing horizon to 5, 10, and 20 days leads to a gradual deterioration in the predicting quality of the index delta. An analysis of the data presented in Figure 3.2.3 suggests that the higher the testing horizon, the weaker the relationship between the risk forecast and its actual realization, right up to its full absence. This conclusion follows both from a comparison of correlation coefficients ($R2 = 0.04$ for 20 days) and from the patterns of points scattering over the regression plane. Besides, the extension of the testing horizon results in the decrease of the slope coefficients of corresponding regression lines. This also supports our observation that the forecasting qualities of the index delta weaken at longer time horizons.

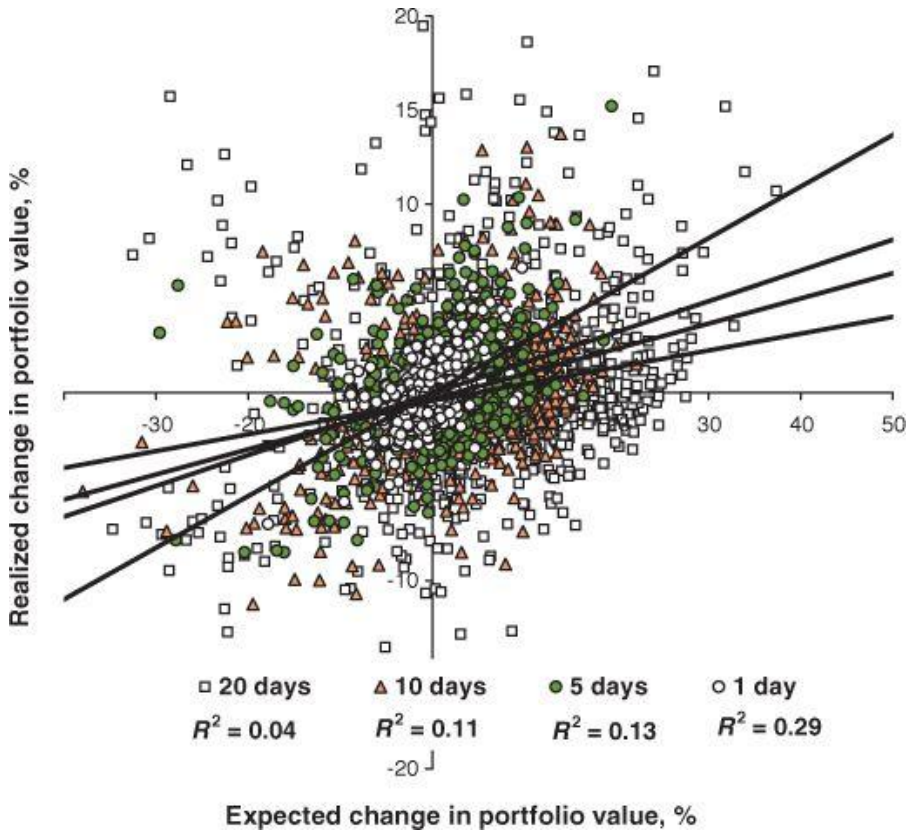

Figure 3.2.3. The relationships between realized and expected (on the basis of index delta) changes in portfolio value. Four testing horizons are shown.

Regression analysis, similar to that presented in Figure 3.2.3, is a simple and intuitively comprehensible tool for qualitative evaluation of the index delta. At the same time, it does not provide a strict

quantitative characteristic for measuring the effectiveness of this risk indicator. For that purpose, it would be preferable to use the difference between the realized and the expected changes in the portfolio value. Since short option positions generate losses when their values increase $(P^{\%}_{realized} > 0)$, positive differences $(P^{\%}_{realized} > P^{\%}_{expected})$ indicate that the risk is underestimated by the index delta. Accordingly, negative differences imply that the risk is overestimated.

Figure 3.2.4 shows the average differences and standard errors (expressing the extent of results variability) for different testing horizons. For the shortest testing horizon (one day), the deviations of the realized changes in the portfolio values from expected ones were close to zero through the whole period, right from portfolio creation up to about 20 days until options expiration. After the 20th day, the closer the portfolios approached the expiration, the more the risk became underestimated. This means that the index delta can forecast risk rather accurately, but for a short period and only at the early stages of the portfolio life (during 30 days from its setup). A 5-day testing horizon gives similar results with the only difference that in this case the underestimation of risk begins earlier (from the 30th day until expiration) and reaches higher values. Further increases in the testing horizon make these tendencies even more pronounced—the underestimation of risk begins earlier and reaches greater extents (see Figure 3.2.4). Besides, it would be worth noting that for short testing horizons, the relationships between the difference and the number of days left until options expiration are nonlinear. At the same time, for the longer testing horizons these relationships gradually become more linear and steeper. This steepness characterizes the speed of the forecast quality deterioration occurring due to the passage of time (approaching the expiration date).

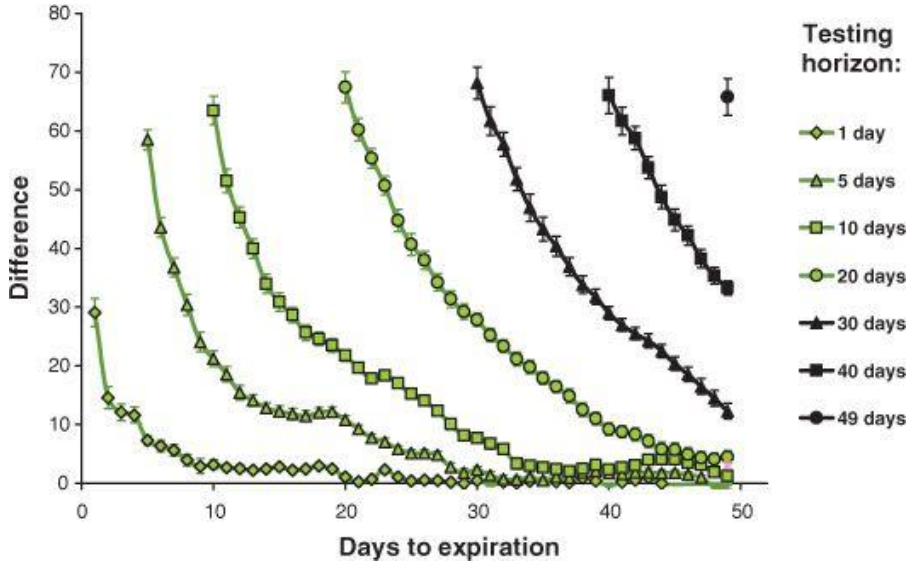

Figure 3.2.4. The dependence of the difference between the realized and the expected changes in the portfolio value on the number of days left until options expiration. Seven testing horizons are shown. Points denote average values; horizontal lines, standard errors.

Complete information on the effectiveness of the index delta and its dependence on the two parameters under examination—the valuation moment (relative to the expiration date) and the testing horizon—can be gathered by presenting the data in the form of a topographic map. In the two-dimensional coordinate system we will plot the number of days left to options expiration (which corresponds to different moments of risk valuation) on the horizontal axis, and the testing horizon on the vertical axis. The height marks of this topographic surface reflect the average difference between the realized and the expected changes in the portfolio value. The topography of such a surface is shown in Figure 3.2.5. The area corresponding to the highest efficiency of the index delta is situated at the bottom-right corner of the surface (highlighted by the broken line on the chart). In general, we can conclude that (1) during the 30 days from the moment of portfolio creation (which covers the period from portfolio creation up to the day when only 20 days are left until the expiration date), the index delta can estimate the risk quite effectively, and (2) the precision of these estimates holds during approximately 25 days from the time when the estimates were made. The triangular shape of the area where the index delta demonstrates its highest efficiency indicates that forecasting the risk for longer periods is possible only at an early stage of the portfolio existence. As the expiration approaches,

the forecast horizon should be shortened.

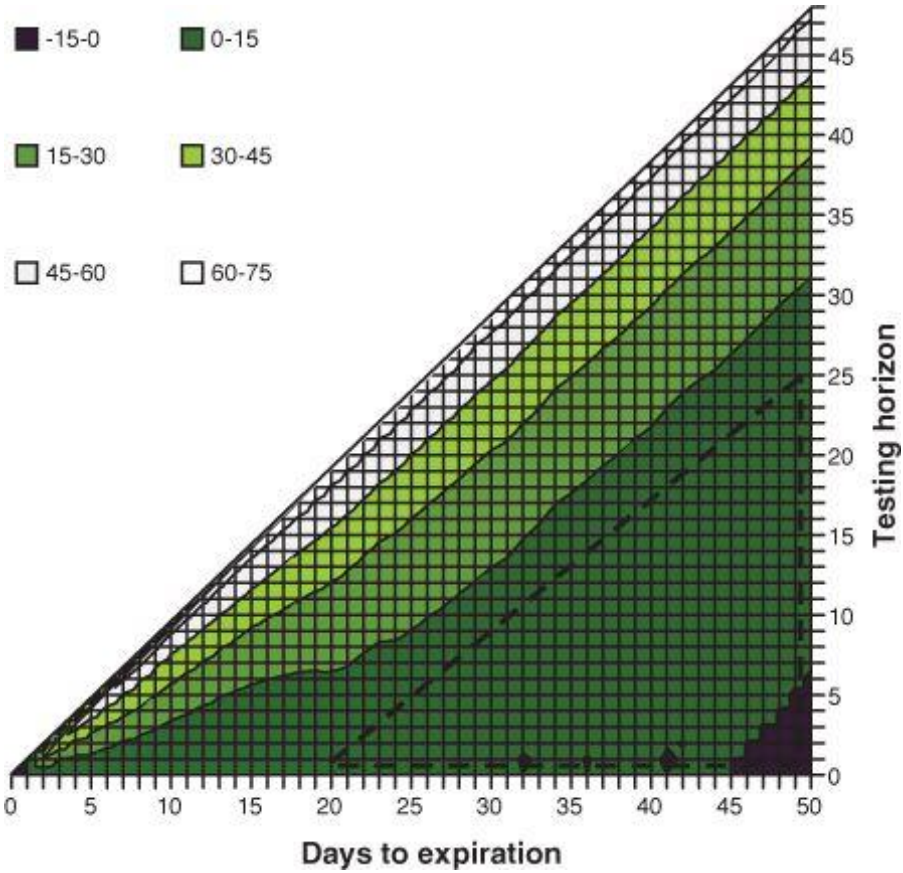

Figure 3.2.5. Dependence of difference between the realized and the expected changes in the portfolio value on the number of days left until options expiration and on the testing horizon. The broken line highlights the area where the risk was estimated by the index delta efficiently.

Applicability of Index Delta

The index delta represents a convenient and easily calculable instrument that can be applied to different types of option portfolios. However, its effectiveness may vary in quite a wide range depending on several factors. In particular, the predicting power of this risk indicator can be affected by the time left to options expiration. When a portfolio is constructed using options with a relatively long expiration date, risk forecasts obtained on the basis of the index delta are quite reliable. Conversely, if a portfolio consists of options that expire shortly, the applicability of the index delta is limited (the risk may turn out to be underestimated significantly).

Besides, we have detected that the effectiveness of forecasting the

risk by the instrumentality of the index delta depends on the magnitude of expected market fluctuations. This indicator demonstrates high forecasting abilities during calm and moderately volatile periods. However, when extreme conditions are prevailing over the markets, the index delta is unsuitable for assessing the risk of option portfolios.

Other things being equal, applying the index delta for risk measurement is quite effective during the initial stage of a portfolio's existence. However, its effectiveness deteriorates as the portfolio ages and the expiration date approaches. Besides, the index delta is able to produce a reliable risk forecast only for relatively short time intervals. There is a direct relationship between the number of days left until options expiration and the forecast horizon—the closer the expiration is, the shorter the forecast should be. Otherwise, the risk may turn out to be underestimated significantly.

3.2.3. Asymmetry Coefficient

This indicator expresses the skewness of the payoff function of the option portfolio. The idea underlying this concept consists in the fact that most strategies based on selling naked options are built on the market-neutrality principle. If the portfolio is really market-neutral, its payoff function should be sufficiently symmetrical relative to the current value of an index reflecting the broad market. Such symmetry implies that the value of the portfolio will change roughly equally regardless of the market direction. If market-neutrality is violated, the payoff function is biased and the asymmetry coefficient can measure this bias.

Since the portfolio value P is the sum of options values that it includes, the relationship between P and changes of the index I can be expressed as

$$P(I,\delta) = \sum_i O_i(A_i(I,\delta)) \cdot x_i, \qquad (3.2.5)$$

where O_i is the value of the option, A_i is the underlying asset, x_i is the quantity of option i in the portfolio, and δ is the index change (for example, $\delta = 0.07$ means that the index rose by 7%). The value of function $A_i(I,\delta)$ can be established using beta β_i, estimated as a slope coefficient in a linear regression between the returns of the underlying asset and the index. If we know beta, we can roughly estimate the value of the underlying asset given that the index changes by a specified amount:

$$A_i(I,\delta) \approx A_i \cdot (1 + \beta_i \cdot \delta). \qquad (3.2.6)$$

By applying the Black-Scholes model, we can estimate the values of

all options included in the portfolio under the assumption that prices of their underlying assets are equal to the values obtained using equation 3.2.6. Summing all O_i values, we get the portfolio value corresponding to equation 3.2.5. Two P values have to be calculated in order to estimate the degree of portfolio skewness—for the case of index growth by $\delta \times I$ and for the case of its decline by the same value. These values will be denoted as $P(A_i(I, \delta+))$ and $P(A_i(I, \delta-))$, respectively. Knowing these two values, the portfolio asymmetry coefficient can be calculated as

$$Asym = \frac{\left|P(A_i(I,\delta^+)) - P(A_i(I,\delta^-))\right|}{2 \cdot I \cdot \delta}. \qquad (3.2.7)$$

If we present the portfolio payoff function by plotting the index values on the horizontal axis and the portfolio values on the vertical axis, then *Asym* can be visualized as the slope of the line connecting the two points with abscissas $X = I \cdot (1 + \delta)$ and $X = I \cdot (1 - \delta)$ and ordinates corresponding to the payoff function. The higher the absolute value of the slope, the more asymmetric the payoff function (if the slope is zero, the payoff function is perfectly symmetric).

Table 3.2.3 contains the interim data required to calculate the asymmetry coefficient for the portfolio consisting of ten short straddles. This portfolio was created on July 21, 2009 (all options expire on August 21, 2009; the risk-free rate is 1%; the quantity of each option is $x_i = 1$). For example, the price of ED stock, provided $\beta = 0.23$ that and that the index rises by 10% ($\delta = 0.1$), is $37.92 \cdot (1 + 0.23 \cdot 0.1) = \$38.79$. Substituting this value into the Black-Scholes formula (instead of the current stock price) gives us the prices of call and put options ($1.89 and $3.06, respectively). Having computed the values of all options in a similar way, we sum them up to obtain $P(A_i(I,\delta+))$ and $P(A_i(I,\delta-))$. The table shows that if the index rises, the portfolio will be worth $46.66, and if it falls, $31.46. Substituting this data into equation 3.2.7 and taking into account that on the date of the portfolio creation the S&P 500 was 954.58, we can calculate the asymmetry coefficient: $Asym = (46.66 - 31.46)/(2 \cdot 954.58 \cdot 0.1) = 0.08$.

Table 3.2.3. Data required to calculate the asymmetry coefficient for the option portfolio.

Stock Ticker	Stock Price	Volatility	Stock Beta	Strike	Option Type	$\delta = 0.1$ Stock Price	$\delta = 0.1$ Option Price	$\delta = -0.1$ Stock Price	$\delta = -0.1$ Option Price
					Call	38.79	1.89	37.05	1.20
ED	37.92	0.49	0.23	40	Put	38.79	3.06	37.05	4.11
					Call	33.90	3.94	29.06	0.22
EIX	31.48	0.15	0.77	30	Put	33.90	0.00	29.06	1.13
					Call	56.90	6.95	49.76	0.60
EXC	53.33	0.11	0.67	50	Put	56.90	0.00	49.76	0.79
					Call	43.26	4.78	39.06	2.26
FE	41.16	0.54	0.51	40	Put	43.26	1.48	39.06	3.16
					Call	63.07	5.56	54.37	1.42
FPL	58.72	0.49	0.74	60	Put	63.07	2.44	54.37	6.99
					Call	13.76	3.78	11.68	1.76
NI	12.72	0.40	0.82	10	Put	13.76	0.00	11.68	0.07
					Call	39.69	0.68	37.31	0.08
PCG	38.50	0.16	0.31	40	Put	39.69	0.94	37.31	2.73
					Call	34.38	4.52	30.42	1.42
PEG	32.40	0.31	0.61	30	Put	34.38	0.12	30.42	0.97
					Call	33.11	3.15	30.75	1.05
SO	31.93	0.15	0.37	30	Put	33.11	0.01	30.75	0.27
					Call	18.33	3.35	14.03	0.14
WMB	16.18	0.26	1.33	15	Put	18.33	0.00	14.03	1.10
Portfolio value P						46.66		31.46	

3.2.4. Loss Probability

As it follows from the name, the loss probability indicator reflects the probability that the portfolio will yield a loss. For a portfolio that contains options, the probability of this negative outcome can be estimated only by a simulation similar to the Monte-Carlo (as described earlier for VaR calculations). To apply this method, a random price should be generated for each underlying asset for a predefined future moment (for example, at the expiration date). The prices are generated on the basis of probability density functions (selected by the system developer) with parameters derived from historical data or predetermined by the system developer according to his personal consideration (scientific method). Then profits and losses of options are calculated for generated stock prices. The sum of these values represents an estimation of the portfolio profit or loss. This cycle represents one iteration. By repeatedly performing many iterations, we can obtain a reliable estimate of the portfolio loss probability.

Table 3.2.4 shows two iterations performed for the portfolio used in the preceding example (see Table 3.2.3). The stock prices were

generated using the lognormal distribution with historical volatility estimated on the basis of a 120-day period (correlations of stock prices were taken into account). The first iteration generated the price of $31.04 for EIX stock, which implies a profit of 30 + 1.74–31.04 = $0.70, while the second iteration performed for the same stock incurred a loss in the amount of $1.28. For the whole portfolio, the first iteration yielded a loss of $2.74, while the second one turned out to be profitable by $5.87.

Table 3.2.4. An example of two iterations performed to estimate the profit or the loss of the options portfolio.

			Iteration 1		Iteration 2	
Stock Ticker	Strike	Straddle Price	Stock Price	Profit	Stock Price	Profit
ED	40	5.08	37.80	2.88	36.03	1.11
EIX	30	1.74	31.04	0.7	33.02	–1.28
EXC	50	3.42	55.38	–1.96	53.22	0.2
FE	40	5.60	31.36	–3.04	35.39	0.99
FPL	60	7.39	68.73	–1.34	67.47	–0.08
NI	10	2.76	11.33	1.43	12.35	0.41
PCG	40	1.99	35.93	–2.08	40.38	1.61
PEG	30	3.17	31.93	1.24	31.30	1.87
SO	30	2.08	32.85	–0.77	31.37	0.71
WMB	15	1.44	16.24	0.2	16.11	0.33
	Total			**–2.74**		**5.87**

A full set of iterations performed for a given portfolio represents a simulation (in the correlation analysis considered in the next section, we generated 20,000 iterations for each simulation). The estimate of the portfolio loss probability is obtained by dividing the number of unprofitable iterations by their total quantity. For example, if 7,420 out of 20,000 iterations used are unprofitable, the loss probability is 0.37 (7,420 / 20,000).

3.3. Interrelationships Between Risk Indicators

An effective risk management system should involve a multitude of alternative indicators based on different principles. Their application area may include initial portfolio construction, assessment, and restructuring of the existing portfolio and generation of stop-loss signals. Different risk indicators should be unique and, as far as possible, interdependent (that is, they should be uncorrelated). This would ensure that each of them supplements the information contained in the other indicators instead of duplicating it. In this

section we examine the interrelationships between the four risk indicators—VaR (the commonly used indicator, applied primarily to linear assets) and three indicators that were developed specifically for assessing the risk of an options portfolio (index delta, asymmetry coefficient, and loss probability).

3.3.1. Method for Testing the Interrelationships

Testing the extent of interrelationships between the four risk indicators is based on the presumption that if different indicators are unique (not duplicating each other), the performances of portfolios created on their basis should be uncorrelated. To test the correlation between the risk indicators, we used the database containing prices of options and their underlying assets from January 2003 until August 2009. For each expiration date, we created 60 series of portfolios (each series was composed of 1,000 portfolios) corresponding to different time intervals left until the expiration. The most distant series was set up 60 days before the expiration, the next one 59 days before it, and so on, right up to the last series. Thus, each number-of-days-to-expiration was represented by 1,000 portfolios, which gives 60,000 portfolios for each expiration day.

Each portfolio consisted of ten short straddles relating to stocks that were randomly selected from the S&P 500 index. Each straddle was constructed using the strike closest to the current stock price. The quantity of options corresponding to each stock was determined as $x_i = 10000 / A_i$, where 10000 represents the equivalent of the capital allocated to each straddle, and A_i is the stock price.

Within each series the values of the four risk indicators were calculated for each of the 1,000 portfolios. Subsequently, the best portfolio was selected for each indicator (thereby, 4 portfolios were chosen from every series). The returns of the selected portfolios were recorded on the expiration date. The returns were expressed in percentage terms and normalized by the time spent in a position (from portfolio creation until the expiration date).

3.3.2. Correlation Analysis

As expected, the returns of portfolios selected on the basis of the four risk indicators were interrelated to a certain extent. However, apart from a single exception, the correlations turned out to be relatively low—in five of the six cases the determination coefficient (squared correlation coefficient) was within the range of 0.3 to 0.4 (see Figure 3.3.1). This implies that information contained in our risk indicators is duplicated by only 30% to 40%. Therefore, the introduction of an additional indicator to the risk evaluation system

that is based on a single criterion can enrich this system with about 60% to 70% of new information. The exception was represented by one pair of indicators—VaR and loss probability. Their high correlation can be interpreted as an indication of the similarity of the basic ideas underlying these indicators. Therefore, their simultaneous application is inexpedient because it will not add a sufficient volume of new information to the whole risk management system.

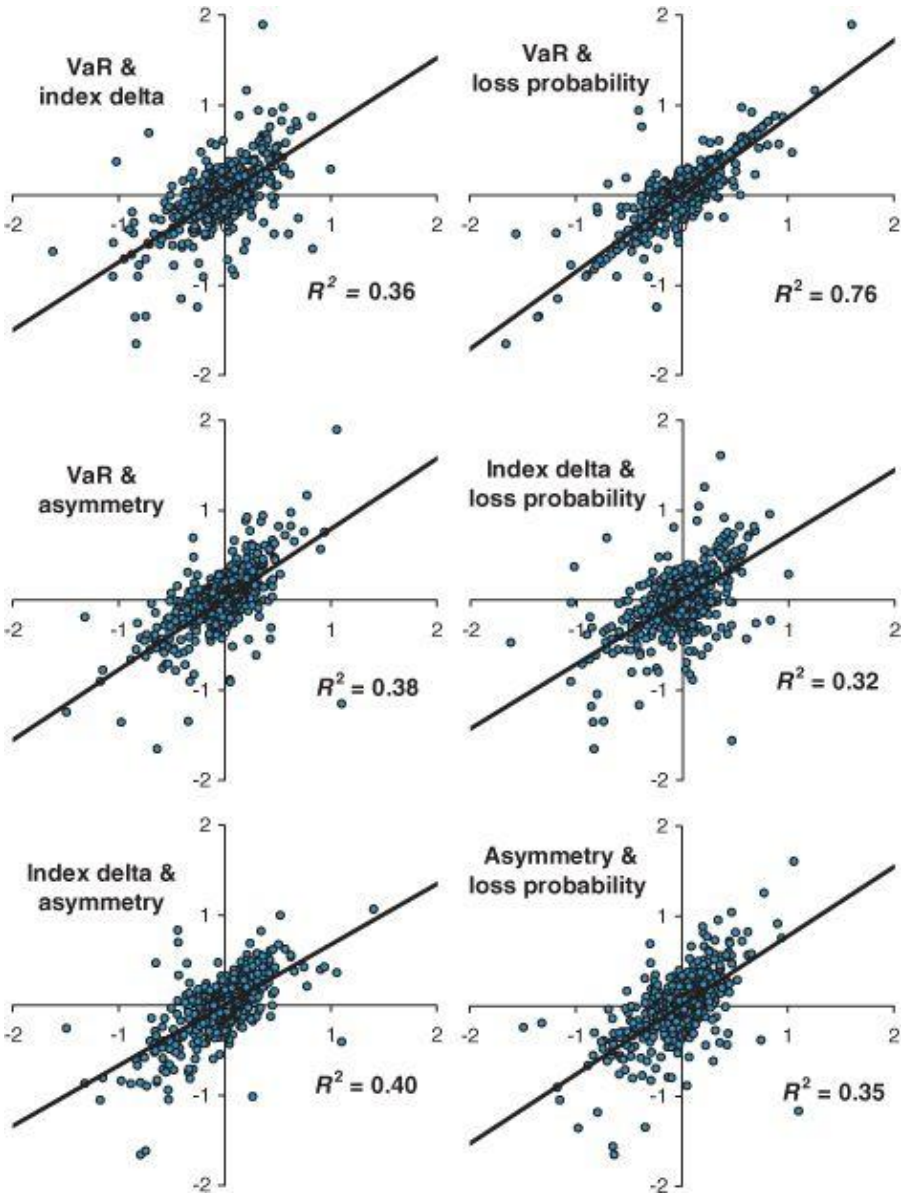

Figure 3.3.1. Correlation of returns of portfolios selected on the basis of the four risk indicators. Each chart shows the pair of indicators and the corresponding determination coefficient.

Let us consider the following issue. Can the degree of correlation between risk indicators vary depending on certain factors? Two factors should be analyzed in the first place: the time interval from the moment of portfolio creation until the options expiration date and the market volatility at the moment of portfolio creation.

Within one day, the extent of the interdependence between the risk indicators can be measured by the variance of returns of the four portfolios selected by these indicators on that day. The higher the variance, the lower the interrelationship between the risk indicators. If the indicators are perfectly correlated, each of them chooses the same portfolio, in which case the variance is zero.

Figure 3.3.2 shows the inverse non-linear relationship between variance and the time interval left until the expiration. This means that close to the expiration date, risk indicators are weakly correlated and, hence, all of them (or, at least, some of them) carry a considerable load of additional information. On the other hand, the variance of returns of the portfolios created long before the expiration is rather low which means that risk indicators are strongly interrelated during this period (hence, the information contained in these indicators overlaps significantly).

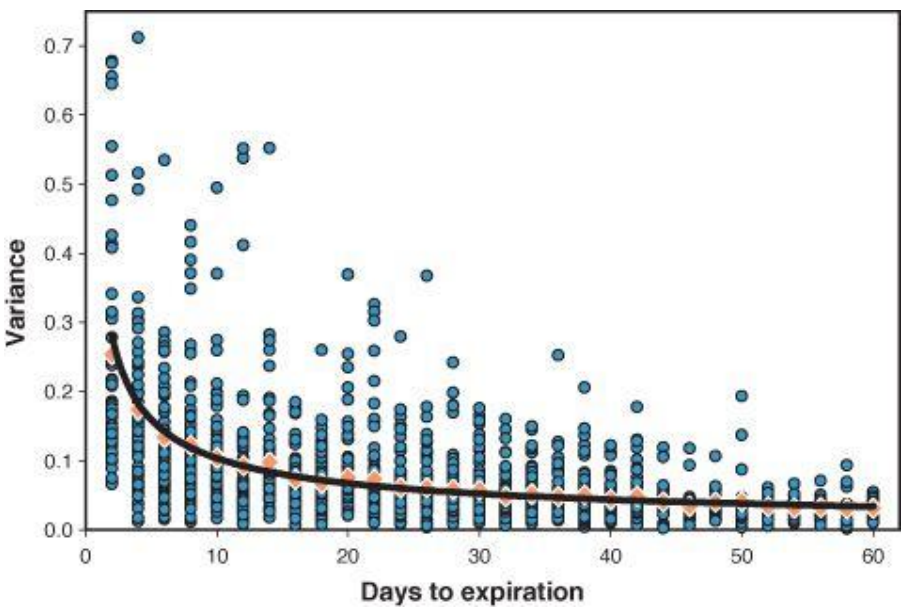

Figure 3.3.2. Relationship between the variance of returns of portfolios selected on the basis of the four risk indicators and the number of days left to options expiration. Higher variance points to lower correlation of risk indicators. Dots denote separate variance values; diamonds denote average values.

Market volatility also influences the interrelationships between the risk indicators, though this effect is much stronger shortly before the expiration than further away from this date. Thus, for portfolios created two days before the expiration, the correlation coefficient between variance and volatility was $r = 0.62$ for historical volatility and $r = 0.68$ for implied volatility (see Figure 3.3.3). Yet, for portfolios created 60 days before expiration, the correlation coefficients were much lower ($r = 0.24$ and $r = 0.28$, respectively). These results indicate that during volatile periods, risk indicators contain additional information, provided that the portfolio is formed of options with a close expiration date. On the other hand, under calm market conditions, different risk indicators carry less nonduplicating information (regardless of the time left to options expiration).

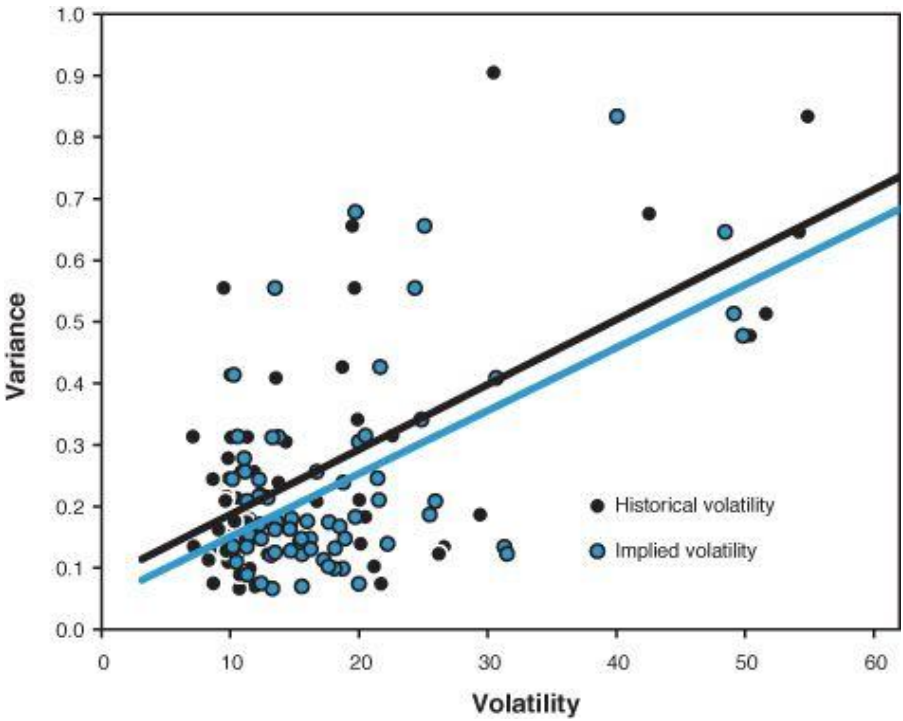

Figure 3.3.3. Relationship between the variance of returns of portfolios selected on the basis of four risk indicators and market volatility (historical and implied). The figure shows only portfolios created two days before options expiration.

3.4. Establishing the Risk Management System: Challenges and Compromises

The set of risk indicators described in this chapter represents an example of an evaluation tool that can be used to develop a

multicriteria risk management system. These four indicators by no means exhaust potential opportunities for creating additional risk forecasting instruments. The efforts in this direction are continuing by both theoreticians and practitioners, which will undoubtedly lead to the development of many useful instruments allowing for a versatile analysis of potential risks threatening option portfolios. The developers of automated trading systems participate actively in this creative process of constructing new and elaborating existing risk management tools. The ever-widening spectrum of available valuation algorithms turns the selection of appropriate indicators that correspond to specific features of the trading strategy under development into a difficult challenge.

This task is further complicated by the fact that the effectiveness of any particular risk indicator may vary depending on many factors. This feature was clearly demonstrated by an example of the index delta, one of the most universal indicators that is suitable for assessing the risk of most options portfolios. The predicting power of the index delta was shown to change depending on the timing of the risk evaluation (relative to the expiration date), on the length of the forecasting horizon, and on the magnitude of expected market fluctuations. Therefore, a wise approach to forming the set of risk indicators for a particular trading strategy consists of including a multitude of them into a risk management system and developing a switching mechanism that allows them to be turned on and off (depending on the favorability of different factors prevailing at that specific moment).

With all that in mind, there is still a necessity to avoid the inclusion of redundant items in the risk evaluation model. This is an important issue, for which a compromise settlement is needed. An effective risk management system should not involve too many indicators. Otherwise, this would overburden calculations and lower the efficiency of the evaluation procedures. While the introduction of any additional indicator into the existing risk management system is being considered, their uniqueness should be thoroughly estimated. All potential candidates must carry really new information not contained in other gauges. Two of the four indicators analyzed in this chapter turned out to duplicate each other: VaR and loss probability. Hence, when creating a trading strategy, founded on the principles that are similar to those used in our examples, the developer should include only one of them in the multicriteria risk management system.

The results of our study suggest that the number of indicators needed for effective risk measuring may change depending on the timing of portfolio creation and on the prevailing market conditions.

In particular, the multicriteria approach to risk evaluation could be the most appropriate one when portfolios are formed close to the expiration date and when the market volatility is relatively high. Under these conditions, evaluation tools are less correlated and, hence, each of them introduces a substantial amount of additional information into the complex risk management system. Undoubtedly, further research will reveal additional factors that influence the extent of interrelationships between different risk indicators and determine the necessity for introducing additional gauges.

Chapter 4. Capital Allocation and Portfolio Construction

Regardless of the type of financial instruments used in the investment process—futures, stocks, currencies, or options—capital management represents an important and complex process. Its significance and influence on trading results is hard to overestimate. The general system of capital management can be viewed at two levels.

The first level represents the distribution of funds among risk-free money market instruments and the risky assets. The principles of creating the first level of a capital management system do not depend on the trading instruments involved and are universal for both options and stock portfolios. There is a multitude of publications on that subject. One of the most useful and complete handbooks on developing the first level of a capital management system is a book by Ralph Vince (Vince, 2007). We will not discuss the details of the first level here.

The *second level* of the capital management system represents the allocation of funds assigned at the first level of the capital management system among separate risky assets. This results in the construction of a certain portfolio. The principles of the second level of capital management are specific for different financial instruments. Therefore, *the development of a capital allocation system requires consideration of many peculiarities that are specific to options.* This is the topic of this chapter.

4.1. Classical Portfolio Theory and Its Applicability to Options

We begin our discussion of the capital allocation procedures with a brief overview of the classical portfolio theory. It will be shown that option portfolio cannot be constructed using the basic form of this fundamental theory. At the same time, its universal principle of balancing risk and return is applicable to all investment assets, including options.

4.1.1. Classical Approach to Portfolio Construction

The problem of optimal capital allocation in the process of portfolio construction arises, since investment in risky assets requires observing the balance between expected return and forecast risks. Capital allocation among different investment assets may lower the total risk

without entailing a significant return decrease. Risk reduction is possible because returns of separate assets may be uncorrelated or may have low correlation. Negative correlations offer even more possibilities of lowering the risk. Besides, if short positions are allowed, correlated assets can also be used to control the risks (long positions are opened in one of the correlated assets; the short position, in another one). Although diversification achieved by allocating capital among different assets inevitably leads to a certain decrease in expected return (as compared to a single asset with the highest expected return), in most cases this effect is overcompensated by a disproportionally greater reduction of risk.

Capital allocation represents a probabilistic problem. In the process of portfolio construction, the investor strives to adapt it to future conditions that have not been realized yet and that can be described only in terms of probability. In most cases the future price dynamics are assumed to follow the same patterns that were observed in the past. Even if this assumption is not made explicitly, probabilities of future price movements are estimated on the basis of past price movements or using parameters that were derived from historical time series. This approach was widely criticized, with numerous examples of events that were realized in spite of their negligibly small probabilities (estimated on the basis of past statistics) presented to support this criticism. Nevertheless, apart from historical time series, there are no other reliable information sources at our disposal for producing probabilistic forecasts (except various expert estimates, which are mostly subjective and hardly formalizable). This problem is solved partially by applying enhanced mathematical models to the construction of probability distributions. In particular, significant efforts are made to substitute lognormal distribution with others describing the dynamics of market returns and probabilities of rare events more precisely.

Modern portfolio theory is based on the classical work by Harry Markowitz. Construction of Markowitz's optimal portfolio is based on the estimations of expected return and risk. All possible variants of capital allocation among the given set of risky assets are considered. Portfolios with the lowest risk for the given return and the highest return for the given level of risk are deemed to be optimal. The complete set of optimal portfolios forms an efficient frontier. The selection of a specific portfolio at this frontier is determined by individual preferences of each investor (their risk aversion, forms, and positioning of indifference curves).

For plain assets with linear payoff functions (stocks, index, or commodity futures), the estimation of return and risk is

straightforward. The expected return of each underlying asset can be estimated on the basis of the risk-free rate, market premium, and beta (CAPM model). Since returns of separate assets are additive, the portfolio return is calculated as the weighted sum of separate returns (weights are equal to shares of capital allocated to each asset). The risk of each underlying asset is expressed through the standard deviation of its price returns calculated on the basis of historical time series. The portfolio risk is calculated analytically using individual standard deviations of each asset, their weights, and the covariance matrix.

4.1.2. Specific Features of Option Portfolios

Classical portfolio theory in its basic form is not applicable to option portfolios. Expected returns of options cannot be estimated on the basis of CAPM or other similar methods. This requires the application of special valuation models. Many of the criteria described by Izraylevich and Tsudikman (2010) can be used as indicators expressing expected return and risk indirectly.

The payoff function of options is not linear. Therefore, the distribution of option returns is not normal and cannot be described properly with the probability density functions of lognormal and most standard distributions. The extent of deviation from the normal distribution in this case is incomparably greater relative to deviations that are usually observed for linear assets. Although in the latter case these deviations are negligible (although many authors convincingly demonstrate that they cannot be neglected, lognormal distribution is still widely used for linear assets), they do cause significant problems when lognormal distribution is applied to options (though the probability density function of this distribution is still used to describe underlying assets returns in options pricing models).

The risk of the option cannot be estimated with the standard deviation (or dispersion) of its price as is done in the classical portfolio theory. Using the standard deviation of the underlying asset price does not solve this problem, since it does not account for specific risks relating to a particular option contract. It is generally accepted to describe risks associated with options in terms of special characteristics called “the Greeks.” The risk of the option price change occurring due to the change in the price of its underlying asset is expressed through *delta*. This characteristic is not additive—it cannot be calculated for the portfolio by summing weighted deltas of options related to different underlying assets. However, using the index delta concept described in Chapter 3, “Risk Management,” can solve this problem.

Another peculiarity of option portfolios relates to the period of options life. Since, unlike with many other assets, the lifetime of any option contract is limited by the expiration date, the historical data necessary to analyze a given option are rather short. Besides, the forecast horizon is limited as well. Whereas for plain assets the expected return is usually estimated on a one-year horizon, for options the forecast horizon cannot exceed the expiration date. At the same time, the limitation of options life has its own advantages. In particular, the validity of fair value estimate has an unambiguous verification date (as opposed to eternal assets for which the date of convergence between estimated fair value and market price cannot be determined objectively).

After the share of each asset is determined, capital allocation within the portfolio consisting of linear assets becomes a trivial task (in contrast to the options portfolio). The only requirement to be observed is that the total investment in all assets is equal to the amount of funds allocated at the first level of the capital management system. When option portfolios are being constructed, the funds assigned at the first level of the capital management system are not invested in full. Opening trading positions in some option combinations requires putting in money (they are referred to as “debit combinations”), other positions result in capital inflows (“credit combinations”), and both may require blocking a certain amount of funds at the trading account (margin requirements).

Margin requirements are the amount of capital necessary to maintain option positions. They are often considered as an equivalent of the investment volume. As a matter of fact, this is not entirely true, since margin requirements may change in time (depending on changing market prices of the underlying assets). Furthermore, there is no standardized algorithm to determine margin requirements. Their calculation methods vary from broker to broker, and even the same broker can change margin requirements depending on market conditions.

The amount of funds assigned at the first level of the capital management system may be used only as a reference point when capital is allocated among portfolio elements. One should aim at two objectives concurrently: Firstly, total margin requirements should not at any point in time exceed this reference amount, and, secondly, assigned capital should be sufficient to fulfill future obligations arising when options are executed or expired.

4.2. Principles of Option Portfolio Construction

In this section we discuss two technical aspects of the capital allocation process. First we examine the number of indicators the developer uses in the process of portfolio construction. Next we consider two alternative approaches to the evaluation of option portfolios for capital allocation purposes.

4.2.1. Dimensionality of the Evaluation System

In the classical portfolio theory capital is allocated on the basis of two indicators: expected return and risk. We suggest considering this two-dimensional system as a special case of a more general approach not limited to a specific number of indicators. There may be more or fewer than two indicators.

One-Dimensional System

Under the one-dimensional system funds assigned at the first level of the capital management system are allocated among portfolio elements proportionally to values of a certain indicator. Such a system may appear to be incomplete, since it is limited to only one indicator —either return or risk estimate. However, for options this approach may be justified, since many indicators used for their evaluation combine forecasts of both return and risk. This can be illustrated by the example of expected profit calculated on the basis of lognormal distribution. It is calculated by integrating the payoff function of an option (or combination of several options) over the probability density function of lognormal (or another) distribution. Two parameters, mean and standard deviation, are necessary to construct the density function. Therefore, this criterion combines forecasts of both return (first parameter) and risk (second parameter).

Certain trading strategies may use a one-dimensional capital allocation system that is not directly related to the evaluation of either return or risk. For example, quantities of different combinations in the portfolio may be determined depending on the amount of premiums received or paid upon their creation. This approach may be appropriate if the trading strategy is based on the idea that all short positions should generate the same amount of premium and all long positions should be equal in price. If the premium of one combination is, for example, $600 and that of the other is $400, the idea is realized by buying (or selling) two contracts of the first combinations and three contracts of the second combination. This method allows opening fewer positions for combinations consisting of more expensive (and hence, in most cases, more risky) options. To some extent this enables the strategy developer to balance portfolio positions in terms of risk.

A one-dimensional system that does not hinge on return and risk estimates may also be based on the principle of stock-equivalency of options positions. This method of capital allocation among combinations included in the portfolio is analogous to the method applied for linear assets (when the amount of funds invested in each asset is equivalent to position volumes). When stock-equivalency of option positions is used for capital allocation, the volume of positions opened for each underlying asset is inversely proportional to its price. Suppose that long positions with the same stock-equivalency value should be established for two options. This means that when options expire, the amount of investment in both underlying assets must be equal (provided that they expire in-the-money). Assume that the strike prices of these two options are $50 and $40. Then we can open 20 contracts (each one consisting of 100 options) for the first option and 25 contracts for the second one. In this example we used a reference amount of $100,000 for both combinations (which is the stock-equivalent amount that we would need to invest in each underlying asset when options are executed). The absolute value of this reference amount may be different and not necessarily the same for all portfolio elements. The total amount required to be invested at options expiration should not exceed the capital allocated to the given position, and the relative ratio of all positions should correspond to the target parameters of the trading strategy. It should be taken into account that using this method is complicated for combinations with different quantities of call and put options. Since it cannot be known in advance which of the options (call or put) will expire in-the-money, the stock-equivalency amount cannot be determined at the moment of position opening. Besides, it is preferable to have coinciding strike prices of put and call options. Failure to comply with this condition causes uncertainty when choosing the strike price to calculate the stock-equivalency value. The calculation of a weighted-average strike price (calculated using all strikes included in the combination) or application of the current price of the underlying asset may be a compromise solution to this problem.

Multidimensional System

The multidimensional evaluation system is based on two or more indicators expressing estimates of return and risk. The usual practice of portfolio construction, originating from the classical work by Markowitz, is based on capital allocation on the basis of two indicators—expected return and standard deviation. However, nothing prevents the strategy developer from introducing additional indicators, if he has sufficient reasons to believe that this will improve the capital allocation system and that potential advantages outweigh

the costs of employing additional computational resources.

When several indicators are used for portfolio construction, the *Pareto method* may be applied in order to identify the best alternative for capital allocation. However, in most cases this results in the selection of numerous portfolios, each of which represents an equally optimal method of capital allocation. The problem of selecting one portfolio from the set of optimal variants is difficult to solve without making subjective judgments. Later we will discuss how to approach this problem.

The alternative approach to implementation of the multidimensional evaluation system consists in using the *convolution of several indicators.* The sum of values of the indicators represents an *additive convolution.* Different weight coefficients may be applied for different indicators. The product of indicators is a *multiplicative convolution.* In this case weights are introduced by raising indicators to a certain power—the higher the power, the higher the importance of the coefficient. Multiplicative convolution is appropriate only for nonnegative values (otherwise, multiplying two negative values would result in positive convolution value).

When calculating the convolution value, it should be taken into account that different indicators can be measured in different units and have different scales. There are several methods of reducing such indicators to a uniform measurement system. In particular, the normalization method can be used for this purpose—the difference between the indicator value and its mean divided by the standard deviation. An alternative method consists in dividing the difference between the indicator value and its minimum by the difference between maximum and minimum values (in this case the criterion value will lie within the interval from zero to one). The former of these methods is more appropriate for the additive convolution; the latter, for the multiplicative convolution.

4.2.2. Evaluation Level

The classical portfolio theory is based on characteristics of return and risk, evaluated for the portfolio as a single whole. This approach —we will call it the *"portfolio approach"*—while being logical and highly reasonable, is still not exhaustive. There are many option strategies for which the capital allocation approach based on the valuation of separate portfolio elements—we will call it the *"elemental approach"*—seems to be more appropriate.

The undisputable advantage of the portfolio approach is the possibility of taking into account the correlations between separate portfolio elements. At the same time, this does not mean that the

elemental approach does not allow taking asset correlations into account. In particular, the average correlation coefficient of a given asset with each of the other assets included in the portfolio may be used as one of the indicators involved in the capital allocation system. Nevertheless, we have to admit that accounting for correlations under the portfolio approach is more proper from the technical and methodological point of view.

Another advantage of the portfolio approach is that there is no need to determine in advance the final set of assets to be included in the portfolio. The portfolio can potentially consist of all available assets, but its specific composition is determined during the process of portfolio formation, when assets with zero weights are excluded from the portfolio. Thus, the subjective judgment in selecting the quantity and the structure of assets in the portfolio is diminished to a great extent. Under the elemental approach all assets satisfying certain conditions (for example, the value of an indicator must exceed the predetermined threshold value) are included in the portfolio. The selection of such indicators and threshold values requires making decisions based largely on subjective estimates.

On the other hand, capital allocation based on the elemental approach may be preferable and, for some trading strategies, the only possible method. In the preceding section we described two examples of capital allocation based on the one-dimensional system of indicators that are not related directly to return and risk estimates. In both cases the capital is allocated using the elemental approach. In the first example, when capital is allocated according to the amount of option premiums, implementation of the portfolio approach is impossible by definition (since capital is allocated by premium amounts relating to each specific combination). The portfolio approach is also inapplicable when capital is allocated according to the principle of equivalency of positions in underlying assets (since the position volume is a characteristic of each separate underlying asset).

Certain valuation criteria that may be used as indicators for the purpose of capital allocation are applicable only at the combination level (since their calculation for the whole portfolio is impossible). For example, accurate calculation of the “IV/HV ratio” for the whole portfolio is unfeasible, since historical volatility relates to a specific underlying asset and implied volatility is a characteristic of a specific option contract. This problem can be solved partially by substituting IV and HV relating to separate options and underlying assets, with market indexes or values derived from these indexes. In particular, IV can be replaced by VIX (index estimating implied volatility of options

on stocks included in the S&P 500 index), whereas HV can be substituted with the historic volatility of the S&P 500 index. However, such substitutions will inevitably reduce the indicator accuracy, unless the portfolio structure matches exactly the employed indexes.

Many indicators used for capital allocation may be applied for both separate combinations and the whole portfolio. In these cases the choice between the elemental and the portfolio approaches is based on peculiarities of each specific trading strategy.

4.3. Indicators Used for Capital Allocation

A great variety of indicators can be developed for capital allocation purposes. Although most of such indicators will express, either directly or indirectly, risk and return characteristics, this is not an indispensable condition. Indicators that are not explicitly related to the estimation of risk and return may turn out to be useful in the process of portfolio construction.

4.3.1. Indicators Unrelated to Return and Risk Evaluation

In section 4.2.1 we discussed, in brief, indicators that are not related to return and risk evaluation. Here we demonstrate the method of portfolio construction using two of these indicators and show that one of them may still (indirectly) express risk.

Capital Allocation by Stock-Equivalency

In the context of capital allocation, we use the notion of equivalent to relate the amount that will be required at the expiration date with the volume of a certain option position. We define the "stock-equivalent of option position" as the amount of capital required at the expiration date or at any other future date, when options are executed. For example, the trading strategy may require that at the expiration date the equal amount of capital be invested in each stock.

To formalize the notion of equivalent, we designate by C a combination consisting of different options on the same underlying asset. Assume that the portfolio is created of m combinations $\{C_1, C_2,..., C_m\}$. Let a_i be the quantity of combination C_i included in the portfolio. If U_i is the price of the underlying asset of combination C_i, then for this combination the equivalent is calculated as product a_iU_i. Let M be the total equivalent of the portfolio (that is, the amount of capital that will be required when all the options included in the portfolio expire). The following equality must hold:

$$M = a_1U_1 + a_2U_2 + ... + a_mU_m. \qquad (4.3.1)$$

This means that the sum of equivalents of all combinations is equal

to the total portfolio equivalent.

Let us consider the example of the trading strategy that uses the capital allocation method requiring an equal amount of capital to be invested in each underlying asset. Suppose that $M = \$1{,}000{,}000$ and that the portfolio consists of 20 short straddles related to stocks belonging to the S&P 500 index (see Table 4.3.1). Each straddle is constructed using one call and one put option and strikes that are nearest to the current stock prices. The portfolio was created on August 28, 2010, using option contracts expiring on September 17, 2010.

Table 4.3.1. Capital allocation among 20 short straddles according to the stock-equivalency and inversely-to-the-premium methods.

				The Number of Combinations If Capital Is Allocated by...	
Ticker	Stock Price	Strike	Premium	Equivalent	Premium
AA	10.01	10	0.74	4,995	3,896
AAPL	240.28	240	14.65	208	195
BA	61.32	62.5	4.14	815	692
CAT	63.95	65	4.26	782	673
CSCO	20.7	21	1.29	2,415	2,220
DD	39.49	39	2.28	1,266	1,259
DELL	11.75	12	0.82	4,255	3,492
EL	56.65	55	3.43	883	836
F	11.17	11	0.91	4,476	3,164
GD	56.83	55	3.55	880	807
GE	14.5	14	1.05	3,448	2,727
GOOG	450.98	450	26.10	111	110
IBM	122.78	125	5.39	407	531
INTC	18.18	18	1.11	2,750	2,591
MA	203.17	200	13.23	246	217
MSFT	23.82	24	1.22	2,099	2,347
ORCL	22.25	22	1.43	2,247	2,002
PG	59.54	60	1.80	840	1,595
T	26.7	27	1.02	1,873	2,821
V	69.91	70	4.24	715	676

The equality of equivalents means that $a_k U_k = a_j U_j$ for each pair of combinations k and j. The following formula follows from equation 4.3.1:

$$a_i = \frac{M}{mU_i}. \quad (4.3.2)$$

Table 4.3.1 shows the position volume for each straddle corresponding to the capital allocation according to the equal equivalent principle. These volumes were calculated subject to satisfying the following two requirements of the strategy: (1) all combinations must have the same stock-equivalent, and (2) the total equivalent of the portfolio equals $1,000,000. For example, for DELL stock, knowing that $m = 20$ and the stock price is $U_7 = 11.75$, the number of straddles is calculated as $a_7 = 1{,}000{,}000 / (20 \times 11.75) = 4{,}255.32$. For the sake of simplicity, we round numbers and ignore lots that are commonly used in real trading.

Allocating Capital Inversely to the Premium

Allocating capital using this method means that the position volume for each option combination is determined on the basis of its premium amount. Let us consider the trading strategy that uses the capital allocation method requiring total premiums received from selling all combinations to be equal. This means that the higher the premium of a given combination, the lower its quantity in the portfolio (that is why this approach is referred to as inverse).

Let p_i and a_i be the premium and the quantity of combination C_i. The equality of premiums in the portfolio $a_j p_j = a_k p_k$ means that numbers $\{a_1, a_2, \ldots, a_m\}$ are inversely proportional to premiums $\{p_1, p_2, \ldots, p_m\}$. Formally, the quantity of each combination in the portfolio is determined as

$$a_i = \frac{M}{p_i \sum_{i=1}^{m} \frac{U_i}{p_i}}. \quad (4.3.3)$$

This method of capital allocation can be illustrated using the same portfolio of combinations as in the preceding example. In Table 4.3.1 the sum of stock price-to-premium ratios is $\sum_{i=1}^{m} \frac{U_i}{p_i} = 349.25$. Combination premiums and their corresponding quantities are shown in the table. For example, the number of combinations for AAPL stock, with the straddle premium $p_2 = \$14.65$, is calculated as $a_2 = 1{,}000{,}000 / (14.65 \times 349.25) = 195.44$.

Comparison of the Two Capital Allocation Methods

Both capital allocation methods have produced proximate results

(see Table 4.3.1). On the one hand, the more expensive the stock, the lower the number of options to be exercised in order to generate the given stock equivalent. On the other hand, the option premium value usually correlates with the stock price, and, thus, for more expensive stocks fewer combinations should be sold to receive the given premium amount.

Although the number of combinations in portfolios created on the basis of two different methods is close, it is not identical. This can be observed in Figure 4.3.1, in which each point represents one of the 20 combinations included in the portfolio. The horizontal axis shows the volumes of positions corresponding to the stock-equivalency method of capital allocation, and the vertical axis shows position volumes calculated on the basis of premiums. If the results of the two methods were the same, all points would lie along the indicated line. However, we can see that points are well scattered around it.

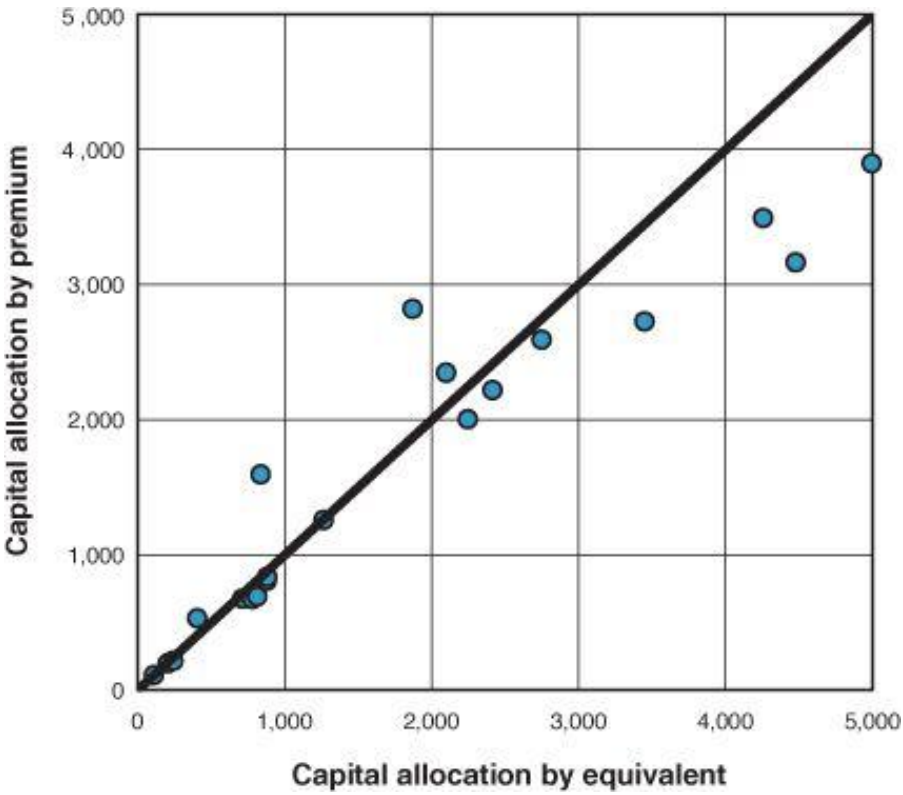

Figure 4.3.1. Relationship between the volume of the position corresponding to the stock-equivalency method of capital allocation (horizontal axis) and the volume of the position calculated on the basis of premium (vertical axis). Each point in the figure represents a specific option combination.

Imperfect correlation between the premium and the stock price is the reason for divergence between two capital allocation methods. The underlying asset price is not the only factor influencing the option premium. One of the main factors determining the option price is the extent of uncertainty regarding the future price of its underlying asset (which is usually expressed through implied volatility). That is why the premiums of two options relating to different stocks with the same price may differ. Hence, other things being equal, the combination with a higher premium is fraught with higher risks and, thus, receives less capital.

This reasoning allows creating an indicator indirectly expressing the extent of combination riskiness. We can state that points situated below the line in Figure 4.3.1 correspond to riskier combinations with a higher premium. This can be expressed through the ratio of the quantity of combinations obtained using formula 4.3.2 to the corresponding quantity obtained using formula 4.3.3. This results in creation of the following risk indicator:

$$riskiness = \frac{a_{equivalent}}{a_{premium}} = \frac{M}{mU_i} \times \frac{p_i \sum_{i=1}^{m} \frac{U_i}{p_i}}{M} = \frac{p_i \sum_{i=1}^{m} \frac{U_i}{p_i}}{mU_i} = \frac{p_i}{U_i} \times \frac{\sum_{i=1}^{m} \frac{U_i}{p_i}}{m} = \frac{p_i}{U_i} \overline{\left(\frac{U}{p}\right)}. \quad (4.3.4)$$

This formula presents risk as the product of the ratio of the *i*-th combination premium to the price of the *i*-th stock and the average ratio of the stock price to the premium. This indicator has a convenient threshold: For riskier combinations (relative to the whole portfolio) its value exceeds one, whereas for less risky combinations it is less than one. If the riskiness of the combination is close to the average riskiness of the portfolio, this indicator converges to one. In the previous examples the average ratio of the underlying asset price to the premium is 17.5. Using data from Table 4.3.1, we can show that for AA stock *riskiness* = (0.74 / 10.01) × 17.5 = 1.28, and for IBM stock *riskiness* = 0.77. This indicates that the former combination is riskier than the latter (and riskier than the portfolio).

The risk indicator calculated using equation 4.3.4 may be used for capital allocation, which ensures that, apart from the premium and the stock price, the riskiness of the combination will also be taken into account. However, we must outline that while this indicator is based on the relative expensiveness of options, it does not evaluate its appropriateness (from the standpoint of historical volatility or anticipated events). Therefore, it cannot be viewed as a comprehensive and all-embracing risk indicator and should rather be used as an auxiliary instrument of capital allocation.

4.3.2. Indicators Related to Return and Risk Evaluation

There are many different indicators expressing estimates of future return and risk in a variety of forms. Here we limit our discussion to two indicators of return (expected profit and profit probability) and three indicators of risk (delta, asymmetry coefficient, and VaR).

In the previous section we estimated the quantity of each combination in the portfolio on the basis of the parameters of a specific combination or its underlying asset. For indicators related to evaluation of return and risk, a more general approach, based on the calculation of weights for all combinations, seems to be more appropriate. To apply such an approach, we need to define a weight function with specified indicator(s) used as argument(s). The value of this function is the weight of each combination in the portfolio.

The weight function can be applied to two types of indicators, which we will call "positive" and "negative." The high values of positive indicators correspond to more attractive combinations; the low values, to less attractive ones. Positive indicators include expected profit and profit probability, as well as all indicators related to forecasting return potential. The high values of negative indicators correspond to a less attractive combination. Most indicators that estimate risk belong to the negative type. For example, the VaR estimating the amount of potential loss has higher values for riskier and, hence, less attractive combinations.

For a given function $\varphi(C)$ the weight of the i-th combination in the portfolio is determined as

$$w_i = \frac{\varphi(C_i)}{\sum_{i=1}^{m} \varphi(C_i)} \qquad (4.3.5)$$

and must satisfy two conditions: $w_i \geq 0$ and $\sum_{i=1}^{m} w_i = 1$.

The method of calculating the quantity of combinations C_i in the portfolio depends on the approach used at the first level of the capital management system. If capital assigned for investment in the option portfolio is the amount that will be required in the future, when options are exercised, then this capital represents the equivalent of portfolio M (see explanations in the preceding section). In this case the quantity of the combinations is calculated as

$$a_i = \frac{M\varphi(C_i)}{\sum_{i=1}^{m} U_i \varphi(C_i)}, \qquad (4.3.6)$$

or with the constant μ:

$$a_i = w_i \cdot \mu, \text{ where } \mu = \frac{M\sum_{i=1}^{m} \varphi(C_i)}{\sum_{i=1}^{m} U_i \varphi(C_i)}. \qquad (4.3.7)$$

If capital F assigned for the investment in the option portfolio represents the amount of investment that is required at the moment of portfolio creation (for example, the total amount of margin requirements), then the quantity of combination C_i is

$$a_i = w_i \cdot F = \frac{F \cdot \varphi(C_i)}{\sum_{i=1}^{m} \varphi(C_i)}. \qquad (4.3.8)$$

In the following sections we will use the approach based on the portfolio equivalent principle (see equation 4.3.7).

Expected Profit and Profit Probability

These two indicators, calculated on the basis of the given distribution (for the sake of simplicity, we will use lognormal distribution in the following examples), represent evaluation criteria that are commonly used for option combinations. Detailed description and calculation algorithms of these criteria can be found in the book by Izraylevich and Tsudikman (2010). Both indicators are positive as their higher values correspond to more attractive combinations. Weight function $\varphi(C)$ for the i-th combination is equal simply to the value of the indicator corresponding to this combination. Table 4.3.2 shows the values of both indicators and weights calculated using equation 4.3.5. Examples in the table involve the same option combinations as in section 4.3.1.

Table 4.3.2. Capital allocation among 20 short straddles using five indicators expressing return and risk estimates.

Ticker	EPLN		PPLN		Characteristic Delta		Asymmetry		VaR		Coefficient of Variation for w_i
	$\varphi(C_i)$	w_i	$\varphi(C_i)$	w_i	$\varphi(C_i)$	w_i	$\varphi(C_i)$	w_i	$\varphi(C_i)$	w_i	
AA	0.0003	0.0025	57.59	0.0457	0.94	0.0560	0.966	0.063	0.933	0.0980	0.69
AAPL	0.0099	0.0826	68.23	0.0541	0.94	0.0560	0.969	0.063	0.063	0.0067	0.56
BA	0.0066	0.0549	61.73	0.0489	0.87	0.0518	0.713	0.046	0.201	0.0211	0.27
CAT	0.0010	0.0085	59.55	0.0472	0.89	0.0534	0.752	0.049	0.164	0.0172	0.41
CSCO	0.0112	0.0935	64.85	0.0514	0.90	0.0539	0.783	0.051	0.761	0.0800	0.39
DD	0.0019	0.0159	59.55	0.0472	0.81	0.0485	0.791	0.051	0.316	0.0331	0.30
DELL	0.0088	0.0732	61.73	0.0489	0.85	0.0511	0.707	0.046	1.089	0.1144	0.58
EL	0.0018	0.0148	58.53	0.0464	0.63	0.0374	0.526	0.034	0.193	0.0203	0.26
F	0.0013	0.0106	57.13	0.0453	0.82	0.0493	0.807	0.052	0.746	0.0784	0.48
GD	0.0070	0.0585	62.43	0.0495	0.60	0.0359	0.493	0.032	0.218	0.0229	0.28
GE	0.0109	0.0907	65.95	0.0523	0.63	0.0375	0.581	0.038	0.782	0.0821	0.50
GOOG	0.0107	0.0892	68.23	0.0541	0.93	0.0553	0.939	0.061	0.038	0.0039	0.61
IBM	0.0060	0.0503	65.38	0.0518	0.78	0.0466	0.609	0.039	0.170	0.0179	0.28
INTC	0.0056	0.0466	62.67	0.0497	0.84	0.0505	0.842	0.055	0.710	0.0746	0.22
MA	0.0045	0.0378	61.73	0.0489	0.79	0.0475	0.764	0.049	0.057	0.0059	0.37
MSFT	0.0061	0.0506	65.57	0.0520	0.96	0.0571	0.869	0.056	0.686	0.0721	0.17
ORCL	0.0191	0.1597	75.56	0.0599	0.84	0.0500	0.818	0.053	0.952	0.1000	0.93
PG	0.0017	0.0144	59.55	0.0472	0.89	0.0533	0.769	0.050	0.417	0.0438	0.31
T	0.0048	0.0401	64.85	0.0514	0.86	0.0515	0.749	0.049	0.863	0.0906	0.39
V	0.0007	0.0057	60.49	0.0480	0.96	0.0575	0.992	0.064	0.161	0.0170	0.52
Σ	0.1196	1	1261.3	1	16.73	1	15.44	1	9.52	1	
Coefficient of Variation		0.80		0.07		0.13		0.18		0.74	

The quantities of all combinations shown in the table were calculated using formula 4.3.7. For example, when capital is allocated by the "expected profit" indicator, the weight and the quantity of the combination relating to CAT stock are estimated as follows. Using stock prices from Table 4.3.1, we calculate $\sum U_i \varphi(C_i) = 11.338$. It follows from Table 4.3.2 that $\sum \varphi(C_i) = 0.1196$. If M = 1,000,000, then μ = 1,000,000 × 0.1196 / 11.338 = 10,550. Considering that for CAT stock $\varphi(C4)$ = 0.001, we obtain the weight $w4$ = 0.001 / 0.1196 = 0.0085 and the quantity of combination $a4$ = 10,550 × 0.0085 = 89.78. When capital is allocated by the "profit probability" indicator, the same algorithm is used to calculate the weight and the quantity of this combination ($w4$ = 0.0472, $a4$ = 580.01).

Delta, Asymmetry, and VaR

Delta expresses the sensitivity of the option price to changes in the underlying asset price. For options relating to the same underlying asset, delta is additive (the delta of the combination is equal to the sum of deltas of separate options included in this combination). The delta of the call option varies from 0 to 1 and the put option delta varies from –1 to 0. Accordingly, the delta of one straddle may range

from –1 to 1. Since the portfolio considered in our examples (see Table 4.3.2) consists of short straddles, the neutrality of combinations to the underlying asset dynamics is a favorable factor. This means that the closer the combination delta to zero, the lower the risk of the position. Hence, for the capital allocation purpose the absolute value of delta should be treated as a negative indicator. The weight function can be defined in the following way: $\varphi(C) = 1 - |\delta(C)|$, where $\delta(C)$ is the delta of combination C.

Asymmetry coefficient is another negative indicator. For a simple combination, like straddle or strangle, it represents the normalized absolute difference between the premiums of call option c and those of put option p:

$$A(C) = \frac{|c - p|}{c + p}.$$

The idea underlying this indicator is the following. Most trading strategies based on shorting of option combinations rely on the expectation that the premium received as a result of options selling will exceed the liabilities arising (at the future expiration date) due to underlying asset price movements. The premium consists of two components: time and intrinsic values. At the moment when a position is entered, the future liability of the option seller equals the intrinsic value (based on an unrealistic, but practically the only possible, assumption that the underlying asset price will not change). Accordingly, the profit potential is higher when the intrinsic value is lower and the time value is higher. This condition is satisfied when the straddle strike approaches the current underlying asset price as closely as possible. The closer the strike price to the current price, the more symmetrical the combination. Since high asymmetry coefficient values are undesirable for short straddles, the weight function can be defined as $\varphi(C) = 1 - A(C)$.

Value-at-Risk (VaR) is the amount of loss that, with a certain probability, will not be exceeded (we will use a 95% probability). In our example (see Table 4.3.2) we use the Monte-Carlo simulation to calculate *VaR* (C) of option combinations. Twenty-thousand paths of future underlying asset movements were simulated (assuming that the underlying asset price is distributed lognormally) for each stock. By substituting the simulated outcomes into the combination payoff function, we have obtained 20,000 payoff variants. Then we discarded 5% of the worst values and selected from the remaining variants an alternative with the lowest payoff. Subtracting the initial combination premium from this value gives an estimated VaR(C) for one combination. Since VaR expresses risk (combinations with lower VaR

values are preferable), capital allocation among combinations should be done inversely proportional to this indicator. The weight function can be defined as $\varphi(C) = 1 / VaR(C)$.

Table 4.3.2 shows weight function values and corresponding weights for the three risk indicators. It also shows variation coefficients (the ratio of the standard deviation to the mean) of weights calculated for each stock and for each indicator. Among stocks, the highest variation coefficient is observed for ORCL (0.93). This means that among all the 20 underlying assets, the amount of capital invested in ORCL combination depends to the greatest extent on the selection of a particular indicator used for capital allocation. Indeed, it follows from Table 4.3.2 that if capital is allocated by expected profit, the share of ORCL in the portfolio is less than 2%. However, if VaR is used to allocate capital, ORCL makes up 10% of the portfolio. The lowest variation coefficient was detected for MSFT stock (0.17). This means that, relative to other underlying assets, the amount of capital invested in MSFT hardly depends on the selection of a particular indicator. Whichever indicator is used to allocate capital, the share of MSFT in the portfolio will vary within a very narrow range (from 5% to 7%).

The comparison of variation coefficients calculated for weights corresponding to each indicator has also given interesting results. Weights obtained on the basis of the profit probability indicator were the least variable (the variation coefficient is 0.07, which is far lower than the coefficients obtained for other indicators). This means that when this indicator is used for capital allocation, each stock receives approximately the same capital share. Hence, profit probability is hardly applicable for capital allocation, at least, in those conditions and for the trading strategy that we used in our example (however, this does not mean that it cannot demonstrate high effectiveness in other circumstances). Delta ranked second in terms of weight variability (0.13), and asymmetry coefficient occupies the third position (0.18). VaR and expected profit have the most variable weights, leading by a huge margin (0.74 and 0.8, respectively). It is interesting that one of the two indicators characterizing return (profit probability) allocates capital most evenly (it has the lowest value of variation coefficient), whereas application of another return indicator (expected return) leads to the most variable capital allocation scenario.

4.4. One-Dimensional System of Capital Allocation

Allocation of capital using a one-dimensional system is relatively simple. This allows the strategy developer to examine in detail

different factors that influence portfolio under construction. Here we demonstrate the effects of three such factors. Next we describe the methodology for measuring the degree of capital concentration within option portfolios and provide several approaches that enable the manipulation of the capital concentration depending on the strategy-specific requirements.

4.4.1. Factors Influencing Capital Allocation

In this section we investigate the influence of various factors on the capital allocation process. While in the previous sections we merely demonstrated the technique of portfolio construction using a limited and predetermined number of combinations, here we consider the long-term functioning of a comprehensive trading system.

The procedure of portfolio construction is simulated for the duration of an eight-year historical period (2002–2010) using the database of stocks and options prices. Stocks included in the S&P 500 index are used as underlying assets for option combinations. Daily close prices are used for stocks and last bids, and asks for options (the average of the spread is considered to be the last option price).

Moving through market history, on each trading day a set of option combinations is created for each stock by applying the following rules. We determine the three nearest option expiration dates. For each expiration date all options on a given stock with strikes distant from the current stock price by no more than 10% are selected. These contracts are used to generate all possible alternatives of short "straddle" and "strangle" combinations (all combinations consist of equal amounts of put and call options;only combinations for which the put strike is less than the call strike are allowed).

As a result, on each successive day a wide set of option combinations is obtained for each of the three nearest expiration dates. For all combinations we calculate the expected profit on the basis of lognormal distribution. Combinations, for which this criterion exceeds 1% of the investment amount, are selected. The portfolio is constructed by allocating $100,000 (the volume of funds assigned at the first level of capital management system) among these combinations. The capital is allocated by one of the seven indicators described in the previous sections:

1. Stock-equivalency of option positions
2. Inverse proportionality to the premium
3. Expected profit
4. Profit probability

5. Delta
6. Asymmetry coefficient
7. VaR

Profit or loss of each portfolio is recorded at the expiration date.

Comparative analysis of these indicators will focus on the following issue. Whether and to what extent do portfolios created on the basis of various indicators differ from each other? To put it in more precise terms, we need to determine the extent to which the portfolio profitability depends on the indicator used to allocate capital among portfolio elements.

In the preceding section we used variation coefficient of combinations' weights to assess the difference in the internal structure of portfolios constructed using different indicators. Since in the examples discussed in the preceding section all weights were positive, using the variation coefficient has not resulted in any problems. However, in this section we compare the various methods of capital allocation by portfolio returns, which may be either positive (if profit is realized) or negative (if loss is incurred). Accordingly, the variation coefficient—which is the ratio of the standard deviation (always positive) to the mean (positive or negative)—may also be negative. Hence, its application to evaluate variability of portfolio returns is impossible. This problem may be solved by expressing return variability through standard deviation not normalized by the mean.

What is this normalization rejection fraught with? It is well known from many practical studies that the standard deviation is positively correlated with the mean. In such cases, trends (or any other patterns) observable in the dynamics of standard deviation may in fact reflect trends of the mean, not of the variability. Normalization eliminates this drawback. Therefore, before beginning our study (in which normalization is inapplicable), we should clarify whether any relationship between the mean and the standard deviation exists. Using nonnormalized standard deviation will be appropriate only if such a relationship is lacking.

The values of the mean and the standard deviation were calculated for each portfolio creation date to assess the relationship between these variables. Returns of seven portfolios, constructed on the basis of seven different indicators, were used as input data for the calculation of the mean and the standard deviation. The regression analysis is shown in Figure 4.4.1. The figure demonstrates no direct relationship between the mean and the standard deviation. Moreover, a slight inverse relationship is observed. Despite the fact that this inverse relationship is statistically significant ($t = 18.4$, $p < 0.001$), we may

neglect it, since the determination coefficient is very low ($R2 = 0.05$). Thus, in our study it would be appropriate to use nonnormalized standard deviation as the measure of variability.

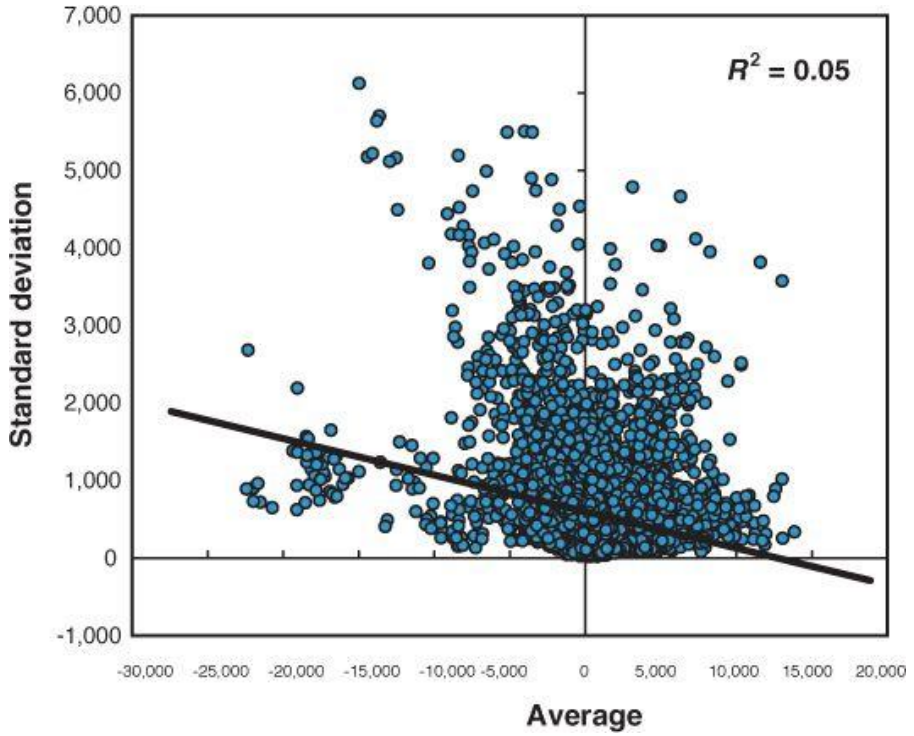

Figure 4.4.1. Relationship between the standard deviation and the mean portfolio return.

During the entire period covered in our study, the level of return variability fluctuated within a quite broad range (see Figure 4.4.2). At certain points in time, the variability was very high. During such periods, the method of capital allocation greatly influenced the portfolio profitability. At the same time, during other periods variability was quite low, which means that the choice of a specific indicator for capital allocation among portfolio elements had an insignificant effect on the portfolio performance.

Thus, we have determined that under some circumstances the choice of a capital allocation method might have a great influence on the trading results, whereas in other conditions this may have absolutely no impact. Therefore, we need to establish the factors that determine the extent of influence exerted by the selection of the capital allocation method on portfolio return. In other words, which factors make the decision about the selection of the specific indicator (used to construct the portfolio) a critical issue? To answer this question, we will examine the influence of three factors (historical

volatility, the timing of portfolio creation, and the number of different underlying assets in the portfolio) on return variability of portfolios differing from each other by the capital allocation method.

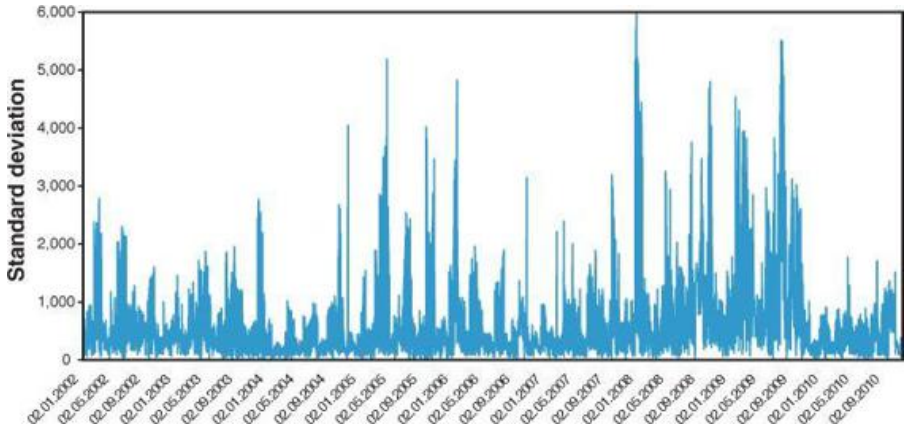

Figure 4.4.2. Time dynamics of return variability (expressed through standard deviation) of portfolios constructed using different indicators.

Historical Volatility

To determine whether market volatility has an influence on the return variability of portfolios created using different indicators, we calculated the historical volatility of the S&P 500 index for each date on which portfolios were constructed (a 120-day historical horizon was used). All return variability values were grouped by the levels of S&P 500 historical volatility. The grouping was performed with the step of 1%. For example, all dates, when historical volatility ranged between 22% and 23%, were included in the same group (and the standard deviations corresponding to these dates were averaged). Accordingly, standard deviations relating to dates when historical volatility ranged between 23% and 24% were included in the next group, and so on.

The results of the analysis are presented in Figure 4.4.3. There is a direct relationship between the volatility level and the return variability of portfolios created on the basis of different indicators. If the market was calm during the period preceding the creation of portfolios (at the time of portfolio creation, the historical volatility was low), the selection of the particular capital allocation method had no significant effect on the strategy profitability (low standard deviation value). On the other hand, if at the moment of portfolio construction the volatility was high, the variability of returns realized at the expiration date was also high. This suggests that when the portfolio is created during periods of volatile market, the choice of the

capital allocation method influences strategy performance significantly.

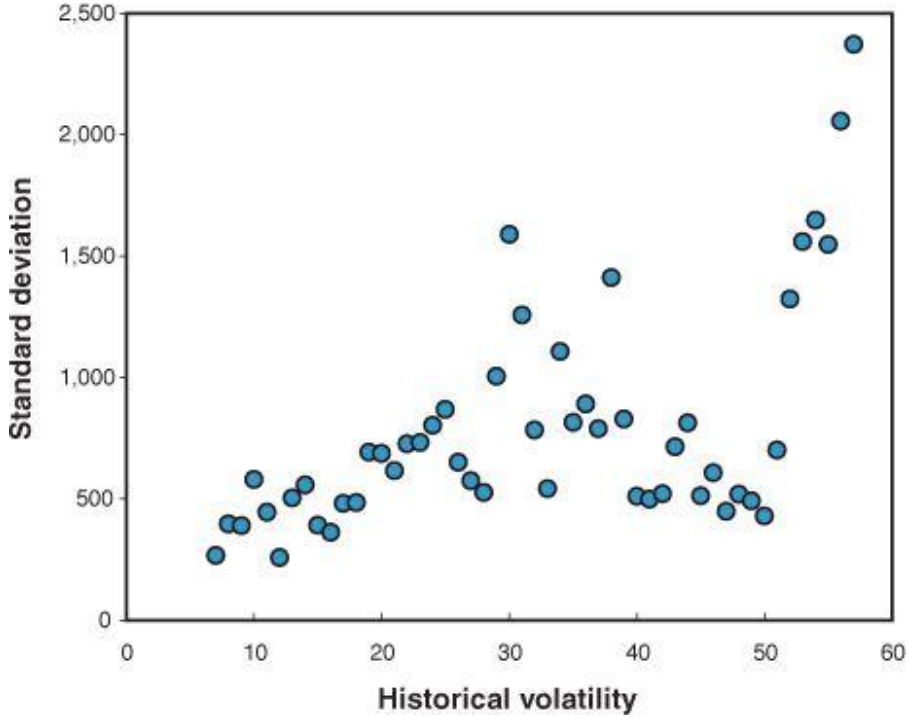

Figure 4.4.3. Relationship between the return variability of portfolios constructed using different capital allocation methods, and the historical market volatility measured at the moment of portfolios creation.

It is worth noting that the pattern of point distribution on the regression plane testifies to the presence of conditional heteroskedasticity in this analysis. This follows from the fact that the dispersion of points (variance) observed at low levels of the independent variable (historical volatility) is much lower than similar dispersion observed at its high levels. This implies that during periods of volatile market return, the variability of portfolios may either be high (as noted previously) or moderate (as follows from the properties of conditional heteroskedasticity).

Number of Days to Options Expiration

Throughout the entire eight-year historical period, seven portfolios were created on each trading day for the nearest expiration date and the two following dates. This allows dividing all portfolios generated in the course of our study into groups according to the number of days from the time of portfolio creation until the expiration date. Averaging of all standard deviations within each group makes it possible to

analyze the relationship between the variability of portfolio returns and the number of days left until options expiration.

Figure 4.4.4 demonstrates that the more days left until options expiration at the moment of portfolio creation, the higher the profit variability of portfolios created using different indicators. If portfolios are constructed shortly before the expiration date, the difference in the profits of portfolios created on the basis of different indicators is negligible. On the other hand, if longer-term options are used to form the portfolio, the selection of a specific capital allocation method influences future profit significantly.

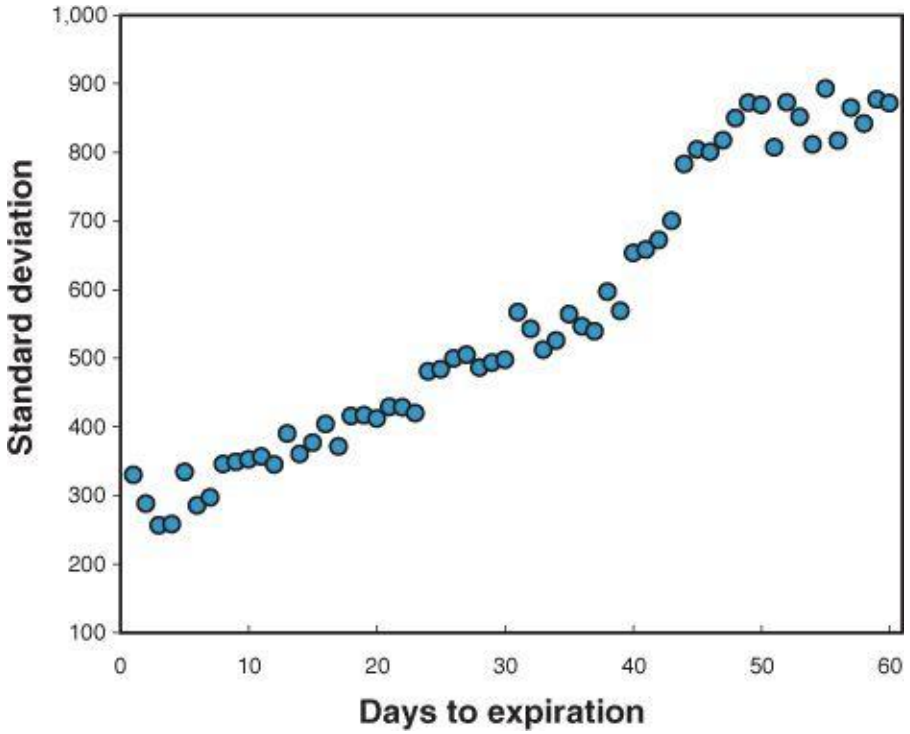

Figure 4.4.4. Relationship between return variability of portfolios constructed using different capital allocation methods, and the number of days left until option expiration.

The practical conclusion following from this observation is important if the trading strategy focuses on the utilization of the nearest option contracts. This type of strategy is quite popular among option traders since the nearest contracts are the most liquid and have the fastest time decay rate (theta). When an automated system that is focused on trading the nearest option contracts is developed, the choice of a particular capital allocation method does not influence final profits. Hence, the developer does not have to spend time and computational resources to search for an optimal capital allocation

system.

Number of Underlying Assets

According to the algorithm of the basic market-neutral strategy, the number of combinations included in the portfolio (and, hence, the quantity of underlying assets to which they relate) is not determined in advance and may vary significantly from case to case. Since the number of different underlying assets reflects diversification level, it would be natural to assume that in the case of poorly diversified portfolios, the selection of a specific capital allocation method would influence future return more than in the case of highly diversified portfolios. Similarly to the previous studies, we grouped all portfolios according to the number of underlying assets. Since the source for creating option combinations that may potentially be included in the portfolios is limited to stocks belonging to the S&P 500 index, the maximum number of underlying assets was 500. Although, from a formal standpoint, the minimum number of underlying assets is 1, in practice none of the portfolios contained a combination related to fewer than 12 stocks.

Data shown in Figure 4.4.5 proves that our assumptions were true. There is a strong inverse relationship between return variability and the number of underlying assets. Furthermore, this relationship is nonlinear. In general, we can state that the higher the number of underlying assets, the lower the profit variability. This means that when the portfolio consists of combinations relating to many different underlying assets, its profit is almost independent of the capital allocation method. However, if the portfolio is not diversified and contains combinations relating to few underlying assets, the capital allocation method has significant influence on future return.

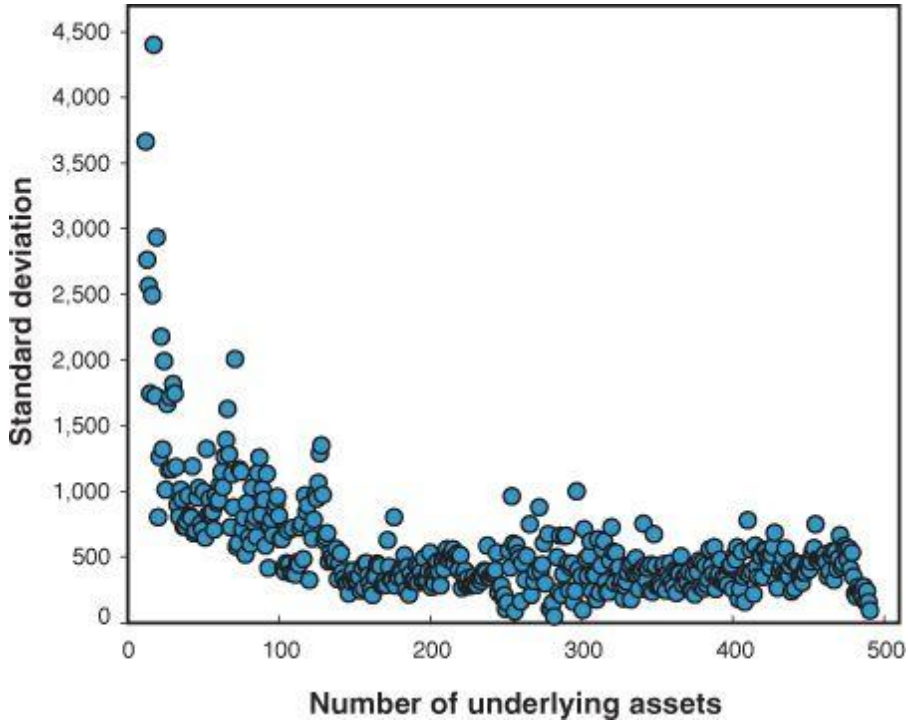

Figure 4.4.5. Relationship between the return variability of portfolios constructed using different capital allocation methods, and the number of underlying assets.

The nonlinearity of the relationship presented in Figure 4.4.5 allows drawing certain quantitative conclusions. When the number of underlying assets increases from dozens to a hundred, profit variability decreases very fast. However, a further increase in the number of underlying assets does not induce any significant variability decrease. This implies that if the trading strategy focuses on creating large complex portfolios containing combinations on 100 or more underlying assets, the time and computational resources consumed in the process of searching for an optimal capital allocation method may be unjustified. On the other hand, when the portfolio consists of combinations relating to a low number of underlying assets, even a small change in their quantity may affect profit drastically (if an inappropriate capital allocation method was selected).

Analysis of Variance

Thus, we determined that the effect of selecting one or another method of capital allocation depends on three factors. So far we have examined (visually) these factors separately. Since in reality they have simultaneous influence on profit variability, we need to perform a statistical test that analyzes the effect of the three factors within the

common framework. This allows verifying whether the relationships that have been detected visually are reliable and makes it possible to give a quantitative expression of the influence exerted by the three factors. The most appropriate statistical tests for this purpose are multiple regression and analysis of variance (ANOVA). These tests are based on the assumption that the relationships between the dependent variable and each of the independent variables are linear. In our case this condition is satisfied for two independent variables (see Figures 4.4.3 and 4.4.4) but not for the third one (see Figure 4.4.5). Hence, before proceeding with the statistical analysis, we need to apply the logarithmic transformation to the "number of underlying assets" variable. This will bring the relationship shown in Figure 4.4.5 closer to linearity.

The results of a multiple regression analysis (see Table 4.4.1) demonstrate that coefficients expressing the influence of the three independent variables (historical volatility, number of days to options expiration, and number of underlying assets) are different from zero at high significance levels. The probabilities that these coefficients diverge from zero by random chance are extremely low (less than 0.1%). The standard approach to interpretation of multiple regressions gives the following formula to be used for forecasting return variability:

Table 4.4.1. Multiple regression and analysis of variance (ANOVA) for the relationship between portfolio return variability (expressed through standard deviation) and three independent variables (historical volatility, the number days until options expiration, and the number of underlying assets in the portfolio).

	Coefficient	Standard Error	*t*-Statistic	Significance
Intercept	893.11	75.26	11.87	< 0.001
Volatility	14.67	0.76	19.22	< 0.001
Log (Number of Stocks)	−376.40	27.12	−13.88	< 0.001
Number of Days	8.29	0.47	17.82	< 0.001
ANOVA	***df***	**SS**	**MSS**	***F***	**Significance**
Regression	3	534,502,733	178,167,578	534.49	< 0.001
Residual	6444	2,148,069,533	333,344		
Total	6447	2,682,572,265			
Multiple R^2	0.20				
Residual Standard Error	577				
Observations	6,448				

$$\textit{Standard Deviation} = 893.11 + 14.67 \times \textit{Volatility} - 376.4 \times \text{Log}(\textit{Stocks}) + 8.29 \times \textit{Days}$$

However, one should be very careful with interpreting and applying this equation. The data shown in Table 4.4.1 suggest that the intercept of the regression is different from zero at a high significance level. This means that, according to the conventional interpretation of the multiple regression, the standard deviation of portfolio return is 893.11 (which is the intercept value), given that all three independent variables have zero values. However, such a conclusion is absurd, because none of these three variables may be equal to zero in the reality. Under such conditions, extrapolation is absolutely impossible. Thus, the analysis presented in Table 4.4.1 should not be used for forecasting the portfolio return variability, but only for revealing the statistical significance of the influence exerted by each of the factors.

The results of ANOVA shown in Table 4.4.1 confirm that the general model of multiple regression is statistically reliable at a high significance level. At the same time, it is worth noting that the value of determination coefficient is rather low ($R2 = 0.2$). This means that all three independent variables (factors affecting the variability of portfolio returns) explain only 20% of the dependent variable dispersion. It would be sensible to assume that another 80% of the dispersion is explained, apart from inevitable random error, by many factors composing the essence of the trading strategy. That is why the relatively small determination coefficient value seems to be reasonable and not surprising.

4.4.2. Measuring the Capital Concentration in the Portfolio

In this section we will compare different indicators by the extent of capital concentration within the portfolio. Suppose that, according to the trading rules of a certain strategy, the amount M has been assigned at the first level of the capital management system for investment in the option portfolio. Also, assume that there are n combinations that can potentially be included in the portfolio. A certain share of capital M must be invested in each of these combinations. In theory, there can be two extreme cases for capital allocation. The whole amount M may be invested in a single combination; all other combinations get no capital (they are not included in the portfolio). Another extreme case consists in even capital allocation, when each combination gets the same portion of funds equal to M/n.

In practice both extreme scenarios are rarely encountered. Usually the capital is allocated in some interim manner, when potentially more attractive combinations receive more capital than less attractive ones. Attractiveness is determined by special indicators, seven of which were discussed in sections 4.3.1 and 4.3.2. Portfolios in which

the bulk of funds is allocated to just a few combinations will be referred to as "concentrated portfolios." Portfolios in which all combinations get approximately the same amount of capital will be called "even portfolios."

The extent of capital concentration is an important characteristic for comparison of different capital allocation methods. Earlier we repeatedly underlined that portfolio diversification level is extremely important for risk management. Up to this point we estimated portfolio diversification by simply counting the number of underlying assets to which combinations included in the portfolio relate (see Figure 4.4.5). However, even if the portfolio consists of combinations relating to a large number of underlying assets, it still can be poorly diversified, if the bulk of funds is concentrated in combinations relating to one (or several) underlying assets. If capital is allocated more or less evenly among combinations relating to different underlying assets, the portfolio is deemed to be more diversified and, hence, less risky.

Earlier we expressed the extent of capital concentration by calculating the variation coefficient for weights of different combinations included in the portfolio (see section 4.3.2 and Table 4.3.2). Since in the current study we need to process an extensive data set (more than 6,000 portfolios for each of the seven capital allocation indicators), it seems reasonable to introduce another indicator, which would be more convenient and statistically sound. We will refer to it as "portfolio concentration index."

We will demonstrate the procedure of concentration index calculation using the data shown in Table 4.3.2. Two indicators will be considered as an example: expected profit and delta. Suppose that two portfolios (each one consisting of the same 20 combinations) were constructed on the basis of these two indicators. These portfolios differ from each other only by the set of weights, w, that are used to allocate capital among combinations. To calculate the concentration index, all combinations are sorted within each portfolio by capital weights. Then we calculate the cumulative proportion of capital invested in two combinations with the highest weights. The same procedure has to be repeated for three combinations, and so on, up to the 20th combination. After that we construct the relationship between the cumulative proportion of capital and the corresponding fraction of combinations used in the calculation of this cumulative proportion (the number of combinations used to calculate the cumulative proportion divided by the total number of combinations in the portfolio). For example, in the case of the portfolio constructed using the "expected profit" indicator, the cumulative proportion

corresponding to the three combinations with highest capital weights is 0.34 (w_{GE} + w_{CSCO} + w_{ORCL} = 0.0907 + 0.0935 + 0.1597 = 0.3439). The fraction of these three combinations in the portfolio is 3 / 20 = 0.15. This means that 34% of the capital is concentrated in 15% of the combinations.

Figure 4.4.6 shows two functions of cumulative capital proportion corresponding to the expected profit and delta indicators. Using these functions, we can calculate the percentage of combinations (out of the total number of combinations in the portfolio) in which *50% of the capital is invested*. This is the value of the portfolio concentration index. For the expected profit indicator the index is 0.25, and for delta it is around 0.42. This means that when capital allocation is based on expected profit, 50% of the capital is invested in 25% of all combinations, and when capital is allocated using delta, half of the funds is invested in 42% of the combinations. Hence, in this example the expected profit indicator leads to more concentrated capital allocation and a less diversified portfolio.

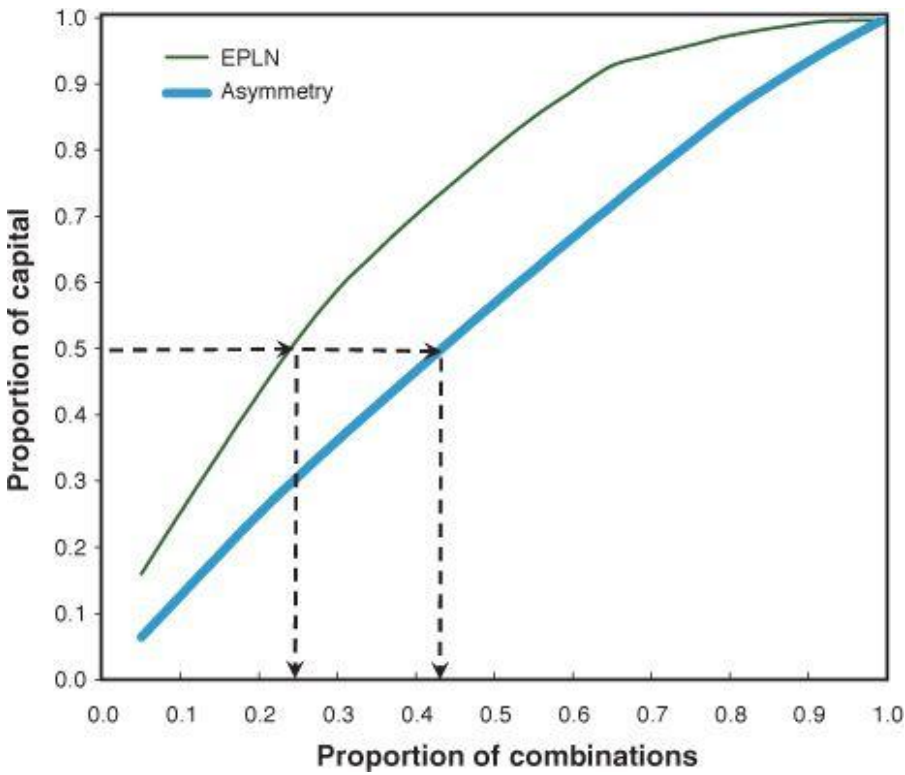

Figure 4.4.6. Visualization of portfolio concentration index calculation using the example of two indicators: expected profit and delta. The data come from Table 4.3.2. Explanations are in the text.

Using this method, we calculated the concentration index for each of the 6,448 portfolios constructed during the eight-year historical period. To compare the extent of capital concentration in portfolios created using different indicators, we built a frequency distribution of the concentration index for each indicator (see Figure 4.4.7).

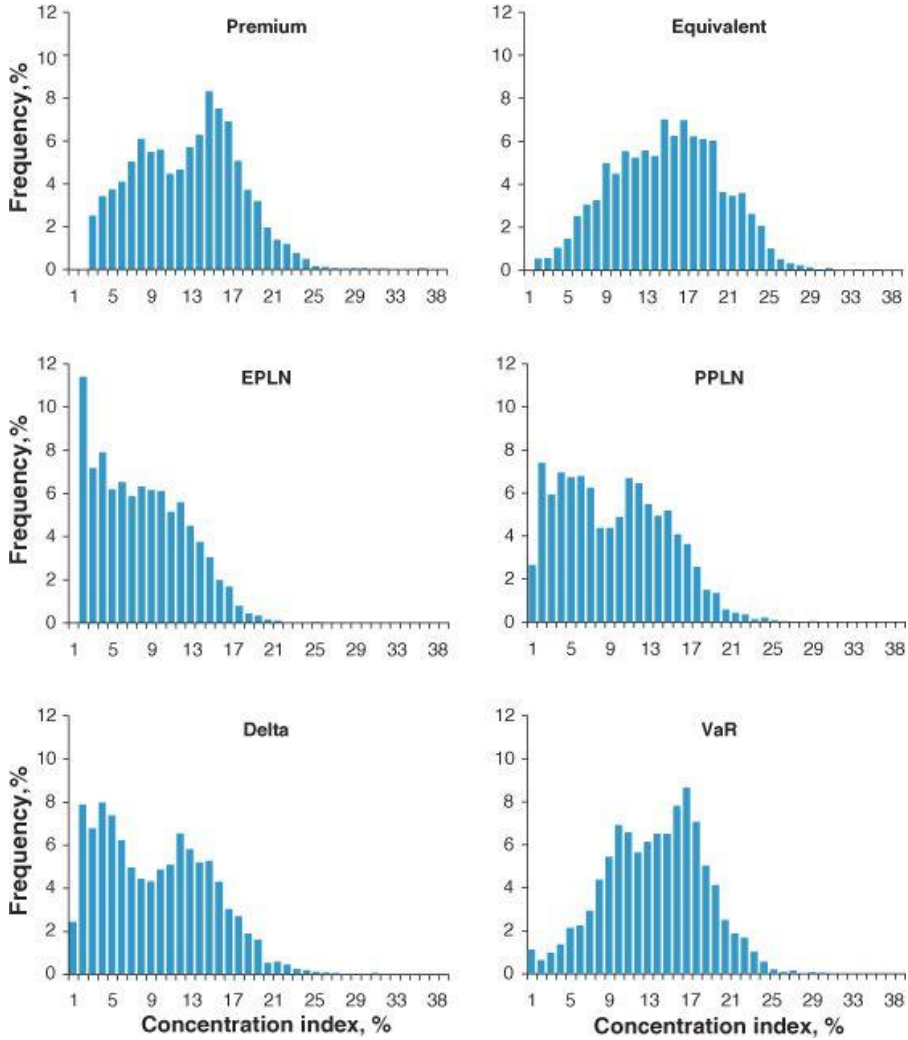

Figure 4.4.7. Frequency distributions of portfolio concentration index for six indicators used to allocate capital among combinations included in the portfolio.

When capital is allocated on the basis of indicators unrelated to the evaluation of return and risk, the shape of the portfolio concentration index distribution is close to normal (the two top charts of Figure 4.4.7). When capital is allocated inversely to the premium, in approximately 8% of the cases half of the funds is invested in 16% to 17% of the combinations. When the portfolio is constructed using the

stock-equivalency indicator, in 7% of the cases half of the capital is concentrated in 15% to 20% of all combinations. Extreme cases, in which 50% of the funds are invested in 1% to 3% of the combinations, were extremely rare (less than 2% of all portfolios).

For portfolios constructed using expected profit, frequency distribution of the concentration index is obviously not normal and resembles the shape of negative binomial distribution (the middle-left chart of Figure 4.4.7). Most often (in 9% to 11% of the cases) the capital is concentrated in 1% to 2% of the combinations. Portfolios with half of the capital invested in more than 15% of all combinations were very rare (less than 4% of all cases).

When portfolios are formed on the basis of the profit probability indicator, the shape of the concentration index distribution approaches uniform distribution (the middle-right chart of Figure 4.4.7). By definition, uniform distribution is characterized by the same frequency of all outcomes. However, in this case the distribution is not completely uniform, since the frequency of concentration index values within the range of 15% and higher is descending.

In those cases when capital in the portfolio is allocated by delta (the bottom-left chart of Figure 4.4.7) and the asymmetry coefficient (not shown on the Figure), the shape of the concentration index distribution resembles the distribution obtained for the profit probability indicator. This indicates the relative evenness of capital allocation among combinations. However, when the portfolio is created on the basis of another indicator related to risk evaluation, VaR, the distribution is close to normal (the bottom-right chart of Figure 4.4.7), which testifies to a lower concentration of capital in the portfolio.

To summarize, we can divide the seven indicators used to allocate capital into three groups (by the extent of capital concentration within the portfolio):

1. Indicators leading to the creation of *highly concentrated portfolios*. In such portfolios a large share of the capital is invested in a relatively small number of combinations. In our study the expected profit represents an example of such an indicator.

2. Indicators leading to the construction of portfolios with *medium capital concentration*. In such portfolios the major portion of the capital is invested in around 15% of combinations. These indicators include VaR, stock-equivalency, and premium-inverse indicator.

3. Indicators leading to the creation of portfolios with *low*

capital concentration (almost even allocation of capital among combinations). In our examples such indicators are represented by delta, profit probability, and asymmetry coefficient.

4.4.3. Transformation of the Weight Function

The weight function $\varphi(C)$ is a function of composite type $\varphi(C) = f(x(C))$, where $x(C)$ is the indicator selected to allocate capital among combinations included in the portfolio. Until now we have assumed that the weight function simply takes on values of the indicator, which is the special case of $\varphi(C) = x(C)$. In such a case the weight of each combination in the portfolio is directly proportional to the value of the indicator corresponding to this combination (graphically the relationship between weight and the indicator is represented by a straight line). However, we are not obliged to be limited to this special case, but should rather allow any type of the relationship between weight function and the indicator. Such relationship should reflect the basic idea of the trading strategy.

For example, the trading system developer may test the alternative when combinations with high indicator values receive more capital than they should get under the proportional (linear) principle of capital allocation. Accordingly, combinations with lower indicator values receive disproportionately less capital. This can be accomplished by transformation of the linear weight function into the *convex function.*

The opposite scenario requires that combinations with high indicator values get less capital than under proportional capital allocation. In this case a disproportionately higher capital share is invested in combinations with low indicator values. To realize such a scenario, the linear weight function should be transformed into the *concave function.*

Different mathematical techniques may be employed to solve this problem. Here we demonstrate the application of the simplest method of transforming the linear function into concave and convex ones. The weight function can be presented as

$$f(x) = y_{min} + (y_{max} - y_{min}) \cdot \left(\frac{x - x_{min}}{x_{max} - x_{min}} \right)^n, \quad y_{min} = f(x_{min}),\ y_{max} = f(x_{max}) \qquad (4.4.1)$$

where x_i is the value of indicator for the i-th combination, x_{min} is the value of the indicator for the combination with the lowest indicator value, and x_{max} is the value of the indicator for the combination with the highest indicator value.

In the further discussion we will assume that $y_{min} = x_{min}$, $y_{max} =$

x_{max}. In this case, if the power value in formula 4.4.1 is 1, the weight function is reduced to a simple linear function. For all $n > 1$ the weight function becomes convex (we will denote it by $f_+(x)$), and for all $0 < n < 1$ this function is concave (we will denote it by $f_-(x)$).

Let us consider an example of calculating the values of convex and concave weight functions (for $n = 2$ and $n = 0.5$, respectively) for the expected profit indicator. We will use the data from Table 4.3.2 that corresponds to the portfolio consisting of 20 option combinations. The minimum value of the indicator is 0.0003, and its maximum value is 0.0191. The value of the convex weight function for AAPL stock is calculated as

$$f^*(x) = 0.0003 + (0.0191 - 0.0003) \cdot \left(\frac{0.0099 - 0.0003}{0.0191 - 0.0003} \right)^2 = 0.0052 .$$

Using the power of $n = 0.5$, we get the value of the concave function for this stock: $f_-(x) = 0.01373$. By applying equation 4.3.5, we can calculate the weight of this combination in the portfolio. If capital is allocated by the convex function, the weight of the AAPL combination is 0.082, and in the case of the concave function it is 0.072.

Having both variants of the weight function calculated in the same manner for all of the 20 stocks, we get two ways of capital allocation —by convex and concave weight functions. The left chart of Figure 4.4.8 shows the two variants of transformed weight functions and the original linear function that was used as the basis for the transformation. The peculiarity of the convex function is that all its values (except for the extremes) are lower than the corresponding values of the basic linear function. For the concave function the opposite is true—all its values (except for the extremes) are higher than those of the basic linear function.

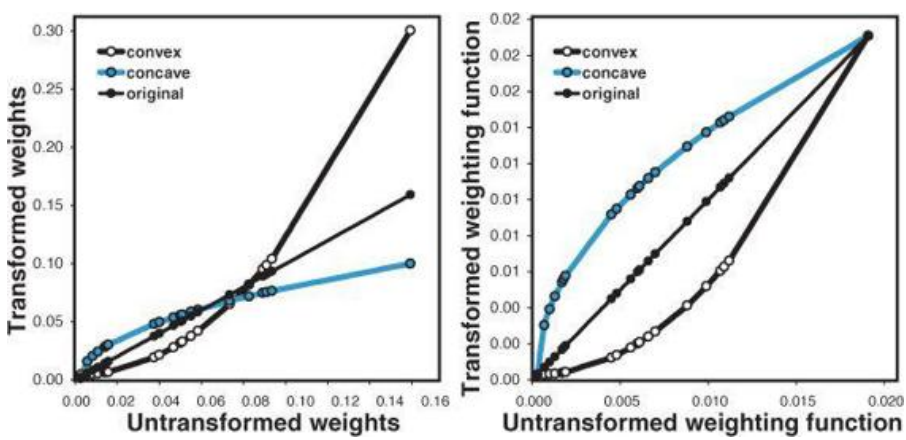

Figure 4.4.8. The left chart shows the relationships between the values of transformed weight functions and their original

(nontransformed) values. The right chart shows the relationships between weights corresponding to transformed functions and the original weights.

The sensitivity of transformed functions to changes in the original weight function is a very important characteristic. Next we describe the properties of convex and concave functions separately for low and high values of their arguments (which are original, nontransformed functions). These descriptions are based on the visual analysis of the left chart of Figure 4.4.8 and on derivatives of function 4.4.1. Assuming $y_{min} = x_{min}$, $y_{max} = x_{max}$, the derivatives of convex ($n = 2$) and concave ($n = 0.5$) functions are

$$\frac{\partial\left(f^{+}(x)\right)}{\partial x} = \frac{2\left(x - x_{\min}\right)}{x_{\max} - x_{\min}}, \qquad \frac{\partial\left(f^{-}(x)\right)}{\partial x} = \frac{0.5\sqrt{x_{\max} - x_{\min}}}{\sqrt{x - x_{\min}}}.$$

Convex Function

At a relatively high level of the original function values, the increment of the convex function is higher than at low levels of the original function values. This means that the difference in transformed function values between the combination with the highest value of indicator and the combination with the next highest value is bigger than the difference in transformed function values between the combination with an average or a low indicator value and the combination with the previous value. Besides, at high levels of the original function values, the increment of convex function values is larger than the increment of the original function itself. Formally this can be expressed as follows. Let $x(C_i)$ be the value of the indicator of the i-th combination. The portfolio consists of m combinations $\{C_1, C_2,\ldots, C_m\}$ and $x(C_m) > x(C_{m-1})$, $x(C_{m-1}) > x(C_{m-2})$ and so on. Then for the convex function the following inequalities hold:

$$f^{+}\left(x\left(C_m\right)\right) - f^{+}\left(x\left(C_{m-1}\right)\right) > f^{+}\left(x\left(C_2\right)\right) - f^{+}\left(x\left(C_1\right)\right),$$

$$f^{+}\left(x\left(C_m\right)\right) - f^{+}\left(x\left(C_{m-1}\right)\right) > x\left(C_m\right) - x\left(C_{m-1}\right).$$

The difference in the values of the transformed function between the combination with the second indicator value and the combination with the first (the lowest) value is smaller than the difference in values of the transformed function between the combination with an average or a high indicator value and the combination with the previous indicator value. Besides, at low levels of the original function values, the increment of the convex function is smaller than the increment of the original function itself. Formally, this is expressed by

the following inequalities:

$$f^{+}\left(x\left(C_{2}\right)\right)-f^{+}\left(x\left(C_{1}\right)\right)<f^{+}\left(x\left(C_{m}\right)\right)-f^{+}\left(x\left(C_{m-1}\right)\right),$$

$$f^{+}\left(x\left(C_{2}\right)\right)-f^{+}\left(x\left(C_{1}\right)\right)<x\left(C_{2}\right)-x\left(C_{1}\right).$$

Concave Function

At high levels of the original function values, the increment of the concave function is smaller than at low levels of the original function values. This means that the difference in values of the transformed function between the combination with the highest indicator value and the combination with the next highest value is smaller than the difference in values of the transformed function between the combination with an average or a low indicator value and the combination with the previous value. Besides, at high levels of the original function the increment of the concave function values is smaller than the increment of the original function itself. Formally, this is expressed as

$$f^{-}\left(x\left(C_{m}\right)\right)-f^{-}\left(x\left(C_{m-1}\right)\right)<f^{-}\left(x\left(C_{2}\right)\right)-f^{-}\left(x\left(C_{1}\right)\right),$$

$$f^{-}\left(x\left(C_{m}\right)\right)-f^{-}\left(x\left(C_{m-1}\right)\right)<x\left(C_{m}\right)-x\left(C_{m-1}\right).$$

The difference in values of the transformed function between the combination with the second indicator value and the combination with the first (the lowest) value is bigger than the difference in the transformed function values between the combination with an average or a high indicator value and the combination with the previous value. Besides, at low levels of the original function, the increment of the concave function is larger than the increment of the initial function itself. Formally, this is expressed by the following inequalities:

$$f^{-}\left(x\left(C_{2}\right)\right)-f^{-}\left(x\left(C_{1}\right)\right)>f^{-}\left(x\left(C_{m}\right)\right)-f^{-}\left(x\left(C_{m-1}\right)\right),$$

$$f^{-}\left(x\left(C_{2}\right)\right)-f^{-}\left(x\left(C_{1}\right)\right)>x\left(C_{2}\right)-x\left(C_{1}\right).$$

Application of Transformed Weight Functions

After all values of the transformed function have been determined, weights are calculated using equation 4.3.5. The right chart of Figure 4.4.8 shows the weights calculated by means of application of the weight functions shown in the left chart of this figure (the row data were taken from Table 4.3.2; the original weight function is expected

profit; the convex and concave functions are calculated using equation 4.4.1 with $n = 2$ and $n = 0.5$, respectively). The straight line on the chart shows weights corresponding to the original (nontransformed) weight function.

It follows from the right chart in Figure 4.4.8 that *when capital is allocated using the convex function,* four combinations with the highest indicator values have higher weights than they would have if the portfolio was created on the basis of the original weight function (these combinations are situated in the interval of high original function values, where the curve of the convex function is above the straight line of the original function). For one of the combinations, the weights obtained using both the convex and the original functions are the same (this combination is situated at the intersection point of the original and the convex function). The other 14 combinations have lower weights under capital allocation based on the convex function than they would have if the capital was allocated using the original weight function (these combinations are situated within the interval of low original function values, where the convex function curve is below the line of the original function).

When *capital is allocated using the concave function,* six combinations with the highest indicator values have lower weights than they would have if the portfolio was constructed on the basis of the original function (these combinations are situated within the interval of high original function values, where the curve of the concave function is below the straight line of the original function). Other combinations have higher weights under the capital allocation scenario based on the concave function than they would have if the portfolio was created using the original function (these combinations are situated within the interval of low original function values, where the concave function curve is above the line of the original function).

These observations lead to the following conclusion: *Portfolios constructed on the basis of the convex function represent a more aggressive approach to capital allocation,* since combinations with higher indicator values receive disproportionately more capital (and combinations with lower values get disproportionately less capital) than they would receive should the portfolio be created using the original weight function. The opposite is true for the concave function, which represents a more conservative approach to capital allocation.

Apart from that, when the convex function is used, capital allocation within the portfolio becomes more concentrated (several combinations receive the bulk of the capital). Using the concave function for portfolio construction leads to more even capital allocation among the portfolio elements. Hence, *portfolios created on*

the basis of the convex function are less diversified than portfolios constructed using the concave function. This also indicates that the first approach is more aggressive than the second one.

Comparison of Convex and Concave Weight Functions by Profit

Using the database encompassing the period of 2002–2010, we will simulate two trading strategies that are similar in all parameters to the strategy described in section 4.4.1, except for the principle of capital allocation. In one case we will allocate capital using the convex function (formula 4.4.1, $n = 2$); in another case, using the concave function (formula 4.4.1, $n = 0.5$). The expected profit will be used as the original weight function.

During the whole simulation period we had created 6,448 portfolios for the convex function and the same number for the concave function. Let us compare profits and losses of portfolios constructed on the basis of these two functions. To begin with, we will analyze the relationship between profit/loss realized when the capital was allocated using one of the transformed weight functions, and profit/loss corresponding to the original weight function.

The pattern of point distribution in the two-dimensional coordinate system (see Figure 4.4.9) demonstrates the effect of transforming the original weight function. To facilitate the interpretation, we added the indifference line with a slope coefficient of 1 to the regression plane. The profit of portfolios situated at this line (each point represents one specific portfolio) is the same regardless of the capital allocation method applied to these portfolios (capital could be allocated using either the original or the transformed weight function without any impact on the portfolio performance). If the profit/loss realized under the capital allocation scenario on the basis of the transformed weight function is shown on the vertical axis, and the profit/loss corresponding to the original function is shown on the horizontal axis, then points situated above the indifference line denote portfolios for which application of the transformed function has led to higher profits or lower losses (as compared to situations when capital was allocated using the original function).

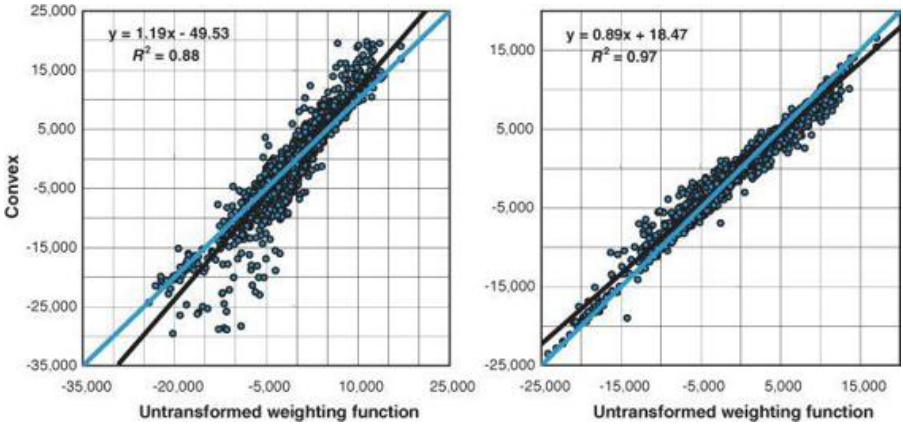

Figure 4.4.9. Relationship between profit/loss of portfolios constructed using the transformed weight function (the left chart, convex function; the right chart, concave function) and profit/loss of portfolios created on the basis of the original (nontransformed) weight function. Each point corresponds to a specific portfolio. The performance of portfolios situated at the light-colored line (slope coefficient = 1) is independent of the capital allocation method. The regression line is black.

When *capital was allocated using the convex weight function* (the left chart in Figure 4.4.9) and the portfolio was profitable (for both the transformed and the original weight functions), the majority of points lie above the indifference line. This means that application of the convex function increases profit. However, in those cases when the portfolio turns out to be unprofitable (for both types of the weight function), the majority of points are situated below the indifference line. This means that when adverse outcomes are realized, losses sustained by portfolios constructed using the convex function exceed losses sustained by portfolios created on the basis of the original function.

Regression analysis enables us to quantitatively express these observations and to verify their statistical reliability. The slope coefficient of the regression line is 1.19, and the slope of the indifference line is by definition 1. Data shown in Table 4.4.2 prove that the difference in slope coefficients is statistically reliable at a high significance level. Therefore, our conclusion that application of the convex function for the capital allocation leads to creating more aggressive portfolios (with higher profit potential and a higher risk of losses) is justified. However, it should be clearly noted that such analysis is appropriate only when the intercept of the regression line is close to zero. In our example, although the intercept is about $50 less than zero and its difference from zero is statistically significant (see

Table 4.4.2), it is still negligibly small relative to the range of values possessed by the variables. Therefore, we can disregard the influence of the intercept and reckon that the profit/loss of portfolios constructed using the convex function is approximately 20% higher than performance of similar portfolios created on the basis of the original weight function (since the slope is 1.19).

Table 4.4.2. Regression analysis of the relationship between portfolio profit/loss realized when capital is allocated using the transformed weight function (either convex or concave function) and profit/loss realized when capital is allocated on the basis of the original weight function.

Transformed Weight Function	Parameters	Coefficient	Standard Error	*t* for Slope Difference from 1	*t* for Intercept Difference from 0	Significance
Convex	Intercept	–49.52836	20.91193		–2.37	< 0.05
	Slope	1.18746	0.00554	33.85		< 0.001
Concave	Intercept	18.47131	7.82187		2.36	< 0.05
	Slope	0.89357	0.00207	–51.39		< 0.001

When *capital was allocated using the concave weight function* (the right chart in Figure 4.4.9), the pattern reverses completely. The majority of portfolios that turned out to be profitable (for both the transformed and the original weight functions) are situated below the indifference line. However, in those cases when portfolios were unprofitable (for both types of the weight function), the majority of points lie above the indifference line. This means that capital allocation on the basis of the concave function leads to a decrease in profit. At the same time, using the concave function diminishes losses occurring when trading outcome happens to be unfavorable.

Although the slope coefficient of the regression line (0.89) is close to 1, it differs from 1 at a high significance level (see Table 4.4.2). Hence, using the concave function to allocate capital leads to creating more conservative portfolios (with lower profit potential and a lower risk of losses).

Comparison of Convex and Concave Weight Functions by Capital Concentration

In the preceding section we compared the profitability of two trading strategies that differ by the shape of the weight functions used to allocate capital among portfolio elements. Now we pass on to a comparison of the same strategies by the extent of capital concentration. Earlier we described the method of calculating the portfolio concentration index and applied it to compare different indicators used in capital allocation (see Figure 4.4.7). The same

method will be applied in the current analysis—the concentration index will be calculated for each of the 6,448 portfolios constructed for each of the two weight functions during the eight-year historical period.

To compare the extent of capital concentration between portfolios constructed on the basis of two transformations of the weight function, we built a frequency distribution of the concentration index. Earlier we demonstrated that when portfolios were created using the nontransformed weight function (expected profit), distribution of the concentration index was not normal and skewed significantly toward low index values (the middle-left chart in Figure 4.4.7). When the convex variant of the weight function was used to allocate capital among combinations, the nonnormality of the frequency distribution became even stronger (see Figure 4.4.10). Most frequently (>16% of all cases), half of the capital was concentrated in just 1% of all combinations. Portfolios with half of their capital invested in more than 15% of the combinations were extremely rare (less than 2% of all cases).

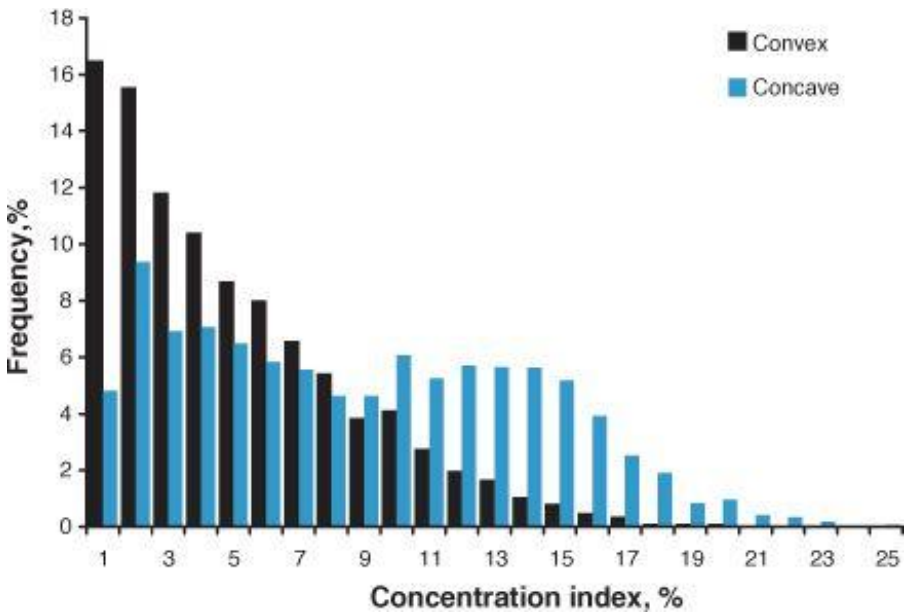

Figure 4.4.10. Frequency distribution of the portfolio concentration index for two variants of the transformed weight function.

Application of the concave weight function for capital allocation among portfolio elements changes the distribution of the capital concentration index fundamentally (compare the middle-left chart in Figure 4.4.7 and the light-colored bars in Figure 4.4.10). In this case

the transformation of the weight function led to almost uniform distribution of the concentration index. Half of the capital is invested in 1% of all combinations, 2% of all combinations, and so on, up to about 18% of all combinations with an approximately equal frequency (4% to 6%).

Our study demonstrates that capital allocation on the basis of the convex function results in the construction of highly concentrated portfolios, in which the bulk of the capital is invested in a few combinations. On the other hand, using the concave weight function leads to the creation of portfolios with more even capital allocation. Since the extent of capital concentration reflects portfolio diversification level, we can conclude that *capital allocation on the basis of the convex function facilitates creation of less diversified and more aggressive portfolios, while using the concave function facilitates creation of more diversified and conservative portfolios.*

4.5. Multidimensional Capital Allocation System

In this section we describe the main methods applicable for creating multidimensional capital allocation systems and compare the properties of portfolio created using one-dimensional system with those of the portfolio that was constructed by the simultaneous application of several indicators.

4.5.1. Method of Multidimensional System Application

The multidimensional system of capital allocation is based on simultaneous application of several indicators expressing estimates of return and risk. The introduction of additional indicators favors creation of a more balanced capital allocation system (from the standpoint of expected return to risks ratio). At the same time, employment of the multidimensional system raises an additional problem that has not been encountered when capital was allocated on the basis of a single indicator—the necessity to select one portfolio out of several equally suitable alternatives. In section 4.2.1 we went through the main approaches to solving this problem. In the following example we apply the multiplicative convolution of several indicators.

Consider capital allocation that is based on two indicators: expected profit and VaR. Calculation of multiplicative convolution of these two indicators and construction of the weight function will be demonstrated using the data from Table 4.3.2. Since these indicators are scaled differently, we need to normalize their values. There are several normalization techniques. We will use the formula reducing values of all indicators to the interval between zero and one:

$$\varphi(C_i)_{\text{normalized}} = \frac{\varphi(C_i) - \varphi(C_{\min})}{\varphi(C_{\max}) - \varphi(C_{\min})}. \quad (4.5.1)$$

Table 4.5.1 shows the original values of indicators expressing expected profit and VaR (the data were taken from Table 4.3.2) and their normalized values calculated by equation 4.5.1. The following example demonstrates the calculation of the normalized expected profit value for AAPL stock. Maximum and minimum values of the expected profit are 0.0191 and 0.0003, respectively. Since the original expected profit of AAPL is 0.0099, the normalized value is calculated as (0.0099 – 0.0003) / (0.0191 – 0.0003) = 0.511.

Since one indicator expresses expected profit and the other one expresses potential loss (VaR), the multiplicative convolution should be calculated as a ratio of expected profit to VaR. Consequently, we face the problem with zero values of normalized indicators. We will solve this problem by substituting zero values with values calculated with the help of the equation

$$\varphi(C_{\min})_{\text{normalized}} = \frac{\varphi(C_{\min})}{\varphi(C_{\min+1})} \cdot \varphi(C_{\min+1})_{\text{normalized}}, \quad (4.5.2)$$

where $\varphi(C_{min+1})$ is the indicator value coming after the minimum one. For example, the normalized function for the expected profit indicator is zero for AA stock. Using equation 4.5.2 and knowing that the stock with the next indicator value is V [$\varphi(C_{min+1}) = 0.0007$], we can calculate the normalized value for AA as $(0.0003/0.0007) \times 0.021 = 0.009$.

After values of all indicators are normalized and convolution values are calculated, we can calculate the weight of each combination in the portfolio. This is done by applying equation 4.3.5 (the results are shown in the last column of Table 4.5.1).

Table 4.5.1. Capital allocation among 20 short straddles using the convolution of two indicators: expected profit and VaR. Normalized values are calculated using equation 4.5.1, except for the values of AA and GOOG stocks (see explanations in the text).

Ticker	Characteristic					
	EPLN		VaR		Convolution	w_i
	$\varphi(C_i)$	$\varphi(C_i)_{\text{normalized}}$	$\varphi(C_i)$	$\varphi(C_i)_{\text{normalized}}$		
AA	0.0003	0.009	0.933	0.852	0.011	0.0001
AAPL	0.0099	0.511	0.063	0.024	21.467	0.2320
BA	0.0066	0.335	0.201	0.155	2.161	0.0234
CAT	0.0010	0.037	0.164	0.120	0.311	0.0034
CSCO	0.0112	0.580	0.761	0.688	0.843	0.0091
DD	0.0019	0.085	0.316	0.265	0.322	0.0035
DELL	0.0088	0.452	1.089	1.000	0.452	0.0049
EL	0.0018	0.080	0.193	0.147	0.541	0.0058
F	0.0013	0.053	0.746	0.674	0.079	0.0009
GD	0.0070	0.356	0.218	0.171	2.081	0.0225
GE	0.0109	0.564	0.782	0.708	0.796	0.0086
GOOG	0.0107	0.553	0.038	0.038	45.900	0.4961
IBM	0.0060	0.303	0.170	0.126	2.414	0.0261
INTC	0.0056	0.282	0.710	0.639	0.441	0.0048
MA	0.0045	0.223	0.057	0.018	12.358	0.1336
MSFT	0.0061	0.309	0.686	0.617	0.500	0.0054
ORCL	0.0191	1.000	0.952	0.870	1.150	0.0124
PG	0.0017	0.074	0.417	0.361	0.207	0.0022
T	0.0048	0.239	0.863	0.785	0.305	0.0033
V	0.0007	0.021	0.161	0.117	0.182	0.0020

4.5.2. Comparison of Multidimensional and One-Dimensional Systems

In this section we analyze the effect of applying the multidimensional capital allocation system. For this purpose we should compare profits/losses of portfolios constructed using a single indicator with profits/losses of portfolios created with simultaneous application of several indicators. The same comparison should be made for the extent of capital concentration within portfolios created using the one-dimensional and multidimensional capital allocation systems.

The comparative analysis was performed during the same eight-year historic period by simulating two trading strategies. Both strategies are similar in all respects to the strategy described in section 4.4.1, except for the capital allocation principle. In one case capital was allocated by the convolution of two indicators (expected profit and VaR); in another case, by the single indicator (expected profit). In total, 6,448 portfolios were created during the entire simulation

period for each of the two capital allocation methods.

Comparison by Profit

By analogy with the study described in section 4.4.3, we analyzed the relationship between the profits/losses realized when the capital was allocated using the convolution of two indicators and the profits/losses realized when the portfolios were constructed on the basis of one indicator. In Figure 4.5.1 profit generated under the capital allocation scenario using convolution is on the vertical axis, and profit generated when the portfolio was created using the only indicator is on the horizontal axis. The profits of portfolios situated at the indifference line (with a slope coefficient of 1) were equal regardless of the capital allocation system employed for portfolio construction. Points located above the indifference line denote portfolios for which application of the multidimensional system has led to higher profits or lower losses (as compared to the case when capital was allocated using the one-dimensional system).

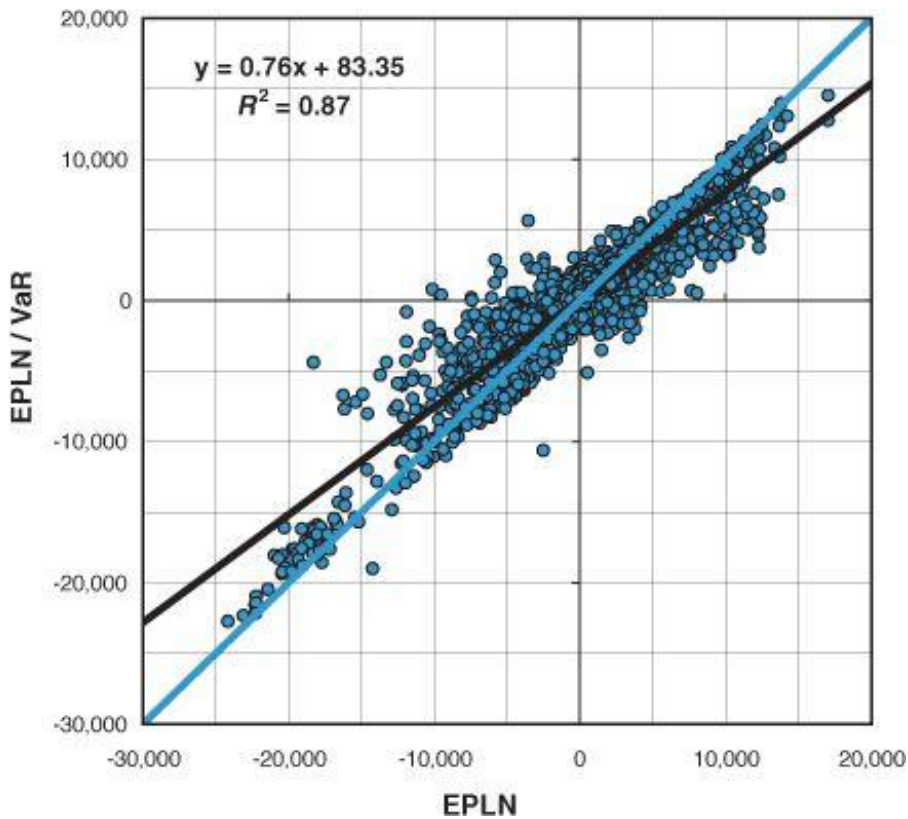

Figure 4.5.1. Relationship between profit/loss of portfolios constructed using convolution of two indicators (expected profit and VaR) and profit/loss of portfolios created on the basis of a single indicator (expected profit). Each point corresponds to a

specific portfolio. Performance of portfolios situated at the light-colored line (slope coefficient = 1) is independent of the capital allocation method. The regression line is black.

In those cases when portfolios were profitable (for both the multidimensional and the one-dimensional capital allocation systems), the majority of points is below the indifference line. This means that introduction of an additional indicator to the system of portfolio construction has led to a decrease in portfolio profitability. At the same time, in those cases when the portfolios turned out to be unprofitable (for both capital allocation systems), the majority of points is situated above the indifference line. This implies that losses of the portfolios created on the basis of the multidimensional capital allocation system were smaller than losses of portfolios constructed using the single indicator.

The regression analysis confirms these observations. The slope coefficient of the regression line is 0.76, which is significantly lower than the slope of the indifference line. Although the intercept is not zero, it is quite small relative to the total range of variables' values. Therefore, its potential impact on the results of our analysis can be neglected. The difference between the slope coefficients of the indifference and the regression lines is statistically reliable at a very high significance level ($t = -64.4$, $p < 0.001$). Thus, we can conclude that *application of the multidimensional capital allocation system leads to creation of more conservative portfolios with lower profit potential and a lower risk of losses.*

Comparison by Capital Concentration

To compare the multidimensional and the one-dimensional systems of capital allocation, we used the method that was described in section 4.4.2. Frequency distributions of the portfolio concentration index were constructed in order to compare the extent of capital concentration within portfolios.

When portfolios were created using a single indicator, frequency distribution of the concentration index was not normal and skewed strongly toward low index values (see Figure 4.5.2). In 9% and 11% of all cases, half of the capital was concentrated in just 1% and 2% of the combinations, respectively. Application of the two-dimensional capital allocation system led to a fundamental change in the concentration index distribution (see Figure 4.5.2). Although the shape of the distribution is quite irregular, its mode shifted significantly toward high index values. In 9% of the cases, half of the capital is concentrated in 16% of all combinations.

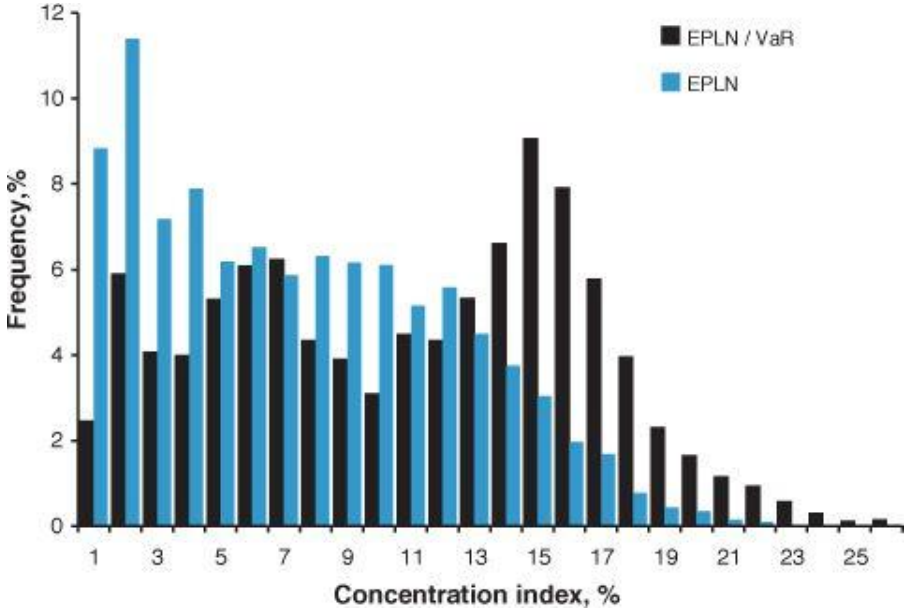

Figure 4.5.2. Frequency distribution of the concentration index for portfolios constructed using the convolution of two indicators (expected profit and VaR) and portfolios created on the basis of the single indicator (expected profit).

Thus, we can conclude that the introduction of an additional indicator to the system of portfolio construction leads to more even capital allocation among option combinations included in the portfolio (as compared to the one-dimensional system). Since the extent of capital concentration reflects the diversification level, we can state that *capital allocation based on the multidimensional system facilitates the creation of more diversified and conservative portfolios.*

4.6. Portfolio System of Capital Allocation

All approaches to capital allocation that were discussed earlier are based on the evaluation of separate elements of the portfolio under construction. In this section we consider another approach to solving the capital allocation problem. This approach is based on the evaluation of return and risk of the whole portfolio rather than separate combinations. The main properties of portfolios constructed using elemental and portfolio approaches are compared.

4.6.1. Specific Features of the Portfolio System

The advantages of the portfolio approach include the possibility to take account of the correlations between separate portfolio elements. The portfolio approach can be applied to both the one-dimensional

capital allocation system (based on a single indicator) and the multidimensional system (based on simultaneous application of several indicators). The classical example of the portfolio approach (as applied to linear assets) implemented in the framework of the two-dimensional capital allocation system is the CAPM model based on indicators of return and risk estimated at the portfolio level.

The fundamentally important property of indicators used for capital allocation under the portfolio system is additivity of their values. This property determines the possibility of calculating the value of a specific indicator for the entire portfolio by summing up its values estimated separately for each portfolio element. We propose the following classification of indicators constructed according to their additivity and applicability at the portfolio level:

- *Additive indicators.* The values of these indicators can be calculated for the whole portfolio by summing their values estimated for each separate asset. The example of such an indicator is profit expected on the basis of lognormal (or any other) distribution.

- *Non-additive indicators that are transformable into additive ones.* Although the values of such indicators cannot be calculated for the portfolio by mere summation, they can be transformed into additive indicators that are inherently similar to the original ones. In the preceding chapter we described the transformation of non-additive delta into the additive index delta.

- *Non-additive indicators that can be calculated analytically.* The values of such indicators cannot be calculated for the whole portfolio by simple summation. Application of special computational techniques coupled with additional information is required to calculate their portfolio values. The example of such an indicator is the standard deviation, which requires (besides standard deviations of separate assets) a covariance matrix (including all portfolio assets) in order to be applied at the portfolio level.

- *Non-additive indicators that cannot be calculated analytically.* The values of such indicators can be calculated for the portfolio of assets neither by mere summation nor by analytical methods. These indicators include various convolutions of several indicators.

When the portfolio system is applied for the purpose of capital allocation, we need to determine such portfolio composition that would maximize the value of an indicator (or a group of indicators)

utilized for portfolio construction. In the case of additive indicators used within the one-dimensional capital allocation system, the solution of this problem is trivial—the total amount should be invested in the single combination with the highest indicator value. Certainly, this solution is inappropriate from the standpoint of diversification. In such a case we can set some minimum weights for all combinations (though this solution may not be satisfactory either). Therefore, it is preferable not to use additive indicators if the portfolio is constructed on the basis of a single indicator.

When capital is allocated on the basis of a non-additive indicator or convolution of several indicators (either additive or non-additive), the maximization problem is usually not solvable by analytical methods. In these cases we have to apply random search methods (for example, the Monte-Carlo method). The maximization problem is then stated as follows: Find such a set of weights for combinations included in the portfolio that maximizes the value of an indicator (or group of indicators) calculated for the portfolio as a single whole.

4.6.2. Comparison of Portfolio and Elemental Systems

In this section we analyze how selection of the evaluation level affects the performance and the characteristics of the portfolio. To perform this analysis, we need to compare profits/losses realized under the capital allocation scenario based on the portfolio system with performance of portfolios constructed on the basis of the elemental system. The same comparison will be arranged for the extent of capital concentration.

The comparative analysis is based on simulation of two trading strategies that are similar to the strategy described in section 4.4.1, except for the capital allocation principle. A total of 6,448 portfolios had been created for each capital allocation method during the entire simulation period. For both strategies the capital was allocated using the convolution of two indicators—expected profit and index delta. For one of the two strategies, the convolution was calculated for each separate combination and the capital was allocated according to the principles of the elemental system (as was done in all examples described earlier). For another strategy the convolution value was calculated for the whole portfolio and capital was allocated using the portfolio system principles. Since values of index delta are directly proportional to the risk of short option combinations, the convolution of the two indicators was calculated as a ratio of expected profit to index delta.

To implement the portfolio system of capital allocation, we need to establish the optimization model. The optimized function is

represented by the convolution of expected profit and index delta calculated for the whole portfolio:

$$\text{Convolution}\,(w_1,\ldots,w_{20}) = \frac{\sum_{i=1}^{n} w_i \cdot \text{EPLN}(C_i)}{\sum_{i=1}^{n} w_i \cdot \text{Index-Delta}(C_i)}, \quad w_i \geq 0, \quad \sum_{i=1}^{n} w_i = 1 \qquad (4.6.1)$$

The solution to the optimization problem consists in finding such a set of weights that maximizes the value of function 4.6.1. This is a nonlinear multiextreme function. The optimization was performed using the Monte-Carlo method by generating random vectors of weights {*w1*,...,*wn*} and selecting the one that maximizes the value of function 4.6.1. The number of iterations (the number of generated random vectors {*w1*,...,*wn*}) was 10,000 for each of the 6,448 portfolios created during the eight-year period.

Comparison by Profits and Losses

Figure 4.6.1 shows the relationship between profits realized when the capital was allocated using the convolution calculated for the whole portfolio and profits realized under capital allocation based on the convolution calculated for separate combinations. Points situated above the indifference line (with a slope coefficient of 1) denote portfolios for which application of the portfolio system of capital allocation has led to higher profits or lower losses (as compared to the case when capital is allocated by the elemental system). Accordingly, points located below the indifference line correspond to portfolios for which application of the portfolio system has resulted in decreased profit or increased losses.

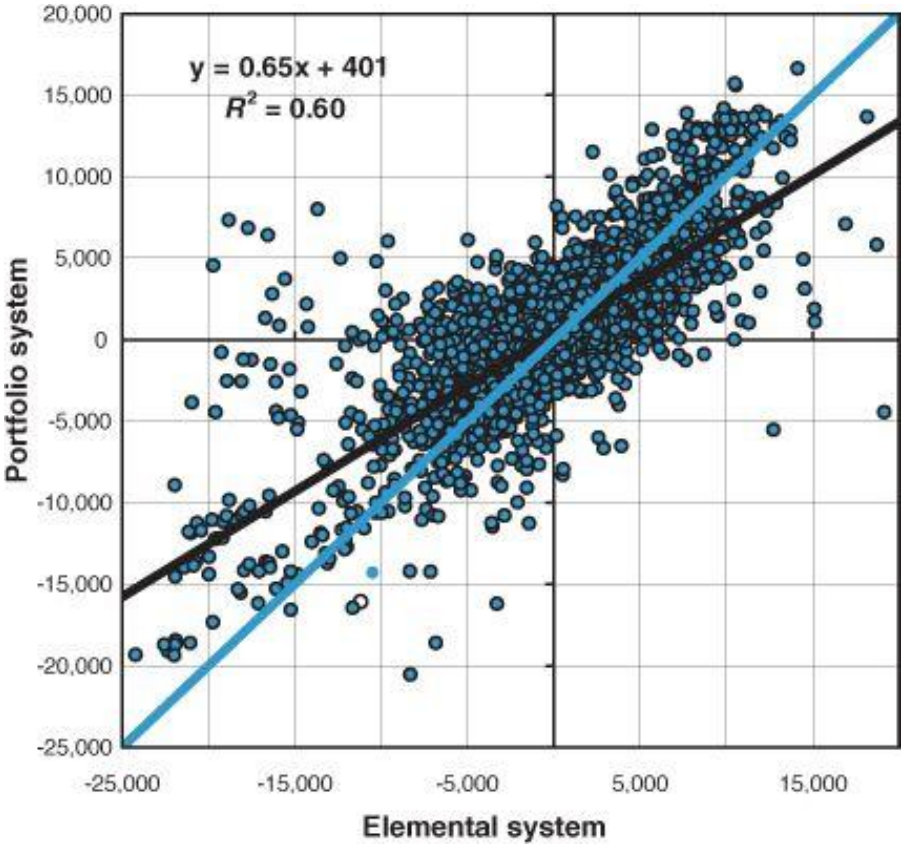

Figure 4.6.1. Relationship between profit/loss of portfolios constructed using the portfolio system of capital allocation and profit/loss of portfolios created on the basis of the elemental system. Each point corresponds to a specific portfolio. Performance of the portfolios situated at the light-colored line (slope coefficient = 1) is independent of the capital allocation method. The regression line is black.

When portfolios turned out to be profitable (under both the elemental and the portfolio capital allocation systems), the majority of points were situated below the indifference line. This indicates that capital allocation based on the evaluation of the whole portfolio decreases profit. However, in those cases when portfolios were unprofitable (under both capital allocation systems), the majority of points were located above the indifference line. This means that portfolios created using the elemental system of capital allocation incurred bigger losses than portfolios constructed using the alternative capital allocation system.

The results of the regression analysis conform to the conclusions drawn from the visual analysis of Figure 4.6.1. The slope coefficient of

the regression line is 0.65, which is much lower than the coefficient of the indifference line. Although in this case the absolute value of intercept is high, it is still quite low relative to the range of variables' values. Therefore, we can neglect its influence on the analysis results (as was the case in the previous studies). The difference of the slope coefficients from 1 is reliable at the high significance level ($t = -53.5$, $p < 0.001$). Thus, we can conclude that *using the portfolio capital allocation system leads to construction of more conservative portfolios with lower profit potential and a lower risk of losses.*

Comparison by Capital Concentration

As in the previous examples, we used the portfolio concentration index described in section 4.4.2 to assess the portfolio and elemental systems. The frequency distribution of the concentration index was used to compare the extent of capital concentration within portfolios that were created on the basis of two different capital allocation systems.

When portfolios were created on the basis of convolution calculated for each separate combination (elemental system), the concentration index distribution was skewed toward the area of low index values (see Figure 4.6.2). The mode of the distribution falls at the concentration index value of 2%. This means that in 10% of all cases, half of the capital was concentrated in just 2% of all combinations included in the portfolio. Application of the portfolio capital allocation system fundamentally changed the shape of the frequency distribution of the portfolio concentration index (see Figure 4.6.2). In this case the distribution was close to normal and its mode was shifted toward higher index values. In about 10% to 12% of all cases, half of the capital was invested in 19% to 22% of the combinations. Extreme cases when half of the capital was allocated in 1% to 5% of all combinations were very rare (less than 1% of the total number of constructed portfolios).

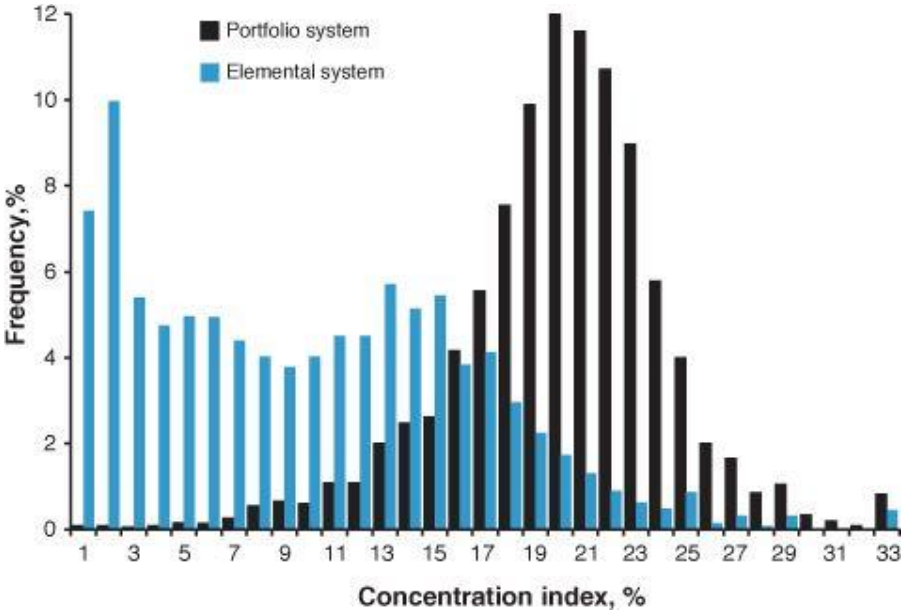

Figure 4.6.2. Frequency distribution of the concentration index for portfolios constructed using the elemental and portfolio capital allocation systems.

Summarizing these observations, we can conclude that relative to the elemental system, application of the portfolio system has led to the creation of portfolios with more even capital allocation. Since the extent of capital concentration reflects the portfolio diversification level, we can state that *capital allocation on the basis of the portfolio system leads to the construction of more diversified and conservative portfolios.*

4.7. Establishing the Capital Allocation System: Challenges and Compromises

It is common knowledge that the choice of capital allocation system can significantly influence the return and the risk of the trading strategy under development. Moreover, improper selection of a capital allocation algorithm can turn a promising strategy into an unprofitable one. As a result, a wonderful trading idea can be rejected at the testing stage.

To create an effective capital allocation system, we need to make a number of compromise decisions. First, we should determine the type of indicators that will be used in the process of portfolio construction. Capital allocation of some option strategies is based on indicators unrelated to return and risk evaluation (although they still indirectly reflect the risk of the created portfolio). For this class of strategies, a

priori fixation of the portfolio diversification level is a priority. We have discussed two algorithms based on this type of indicator. One of them allocates capital on the basis of the option premiums received or paid when positions are opened, and the other one is based on the estimations of the stock-equivalency of option positions.

However, the capital allocation system of most trading strategies is based on indicators that directly express return and risk estimates. Selection of a specific indicator or group of indicators depends on the unique features of the developed strategy and its basic ideas. We can literally state that this is the most critical stage of the capital allocation algorithm development process. Although in this chapter we have discussed only five indicators of this type, in practice there may be many more of them. Furthermore, each indicator may have many variations determined by the selection of specific parameter values and modifications of different functions incorporated in the computational algorithm. We demonstrated that the choice of a particular indicator influences the portfolio performance significantly. At the same time, we have determined three factors that make the trading strategy even more sensitive to appropriate selection of indicators. In particular, we found out that during periods of high market volatility the profitability of the strategy shows greater dependency on the capital allocation method than during calm market periods. If option combinations included in the portfolio relate to a small number of underlying assets and if these combinations consist mainly of long-term options, the selection of specific indicators becomes more critical than when portfolios are more diversified and constructed using shorter-term options.

After a specific indicator (or a set of indicators) has been selected, the extent of its influence on the capital allocation system can be modified. We have proposed the procedure of indicator transformation (based on the convex and concave weight functions) that leads to construction of more (or less) aggressive portfolios, depending on the preferences of the trading strategy developer. Using this procedure, the developer can adjust the strategy to a desired return level, and he can also adapt it to the parameters determined by the risk aversion of a specific user.

The next step in developing an algorithm of portfolio construction is the decision concerning the dimensionality of the capital allocation system. The simplest one-dimensional system based on a single indicator may be rather satisfactory. However, in many cases we have to use at least the two-dimensional system, in which one of the indicators expresses the return and the other indicator estimates risk. In this chapter we have demonstrated that introduction of an

additional indicator into the capital allocation system may lead to the creation of more conservative portfolios with lower return and risk. However, these results should be treated with caution (only as indicative suggestions), since using other indicators (or application of the same indicators to another strategy) may lead to different conclusions. In any case, defining the dimensionality of the capital allocation system is a trade-off between potential advantages of additional indicators and their costs consisting of increased consumption of computational resources and time required to make calculations.

Along with the selection of specific indicators, another critical decision made by the strategy developer in the course of creating the capital allocation system is the choice of an evaluation level. The capital may be allocated on the basis of estimates performed for the whole portfolio. The alternative approach is to estimate separate elements of the portfolio under construction. Our study shows that evaluations executed at the portfolio level lead to creation of less aggressive and more diversified portfolios with lower expected profits that are, however, compensated for by smaller potential losses.

Chapter 5. Backtesting of Option Trading Strategies

An automated trading system represents a package of software modules performing functions of development, formalization, setting up, and testing of trading strategies. Development and formalization were discussed in Chapter 1. Setup of strategies, which includes parameters optimization, capital allocation, and risk management, was described in Chapters 2, 3, and 4, respectively. The current chapter is dedicated to testing options trading strategies on historical data (such testing is commonly referred to as "backtesting").

The module performing backtesting functions is composed of several objects: a historical database, a generator of position opening and closing signals, a simulator of trading orders execution, and a system of performance evaluation. These components are used to model a long-term trading process and to evaluate its results.

Backtesting of trading strategies has received wide coverage in financial literature. There is a large corpus of works written at different levels of complexity and elaboration. The majority of these publications describe the backtesting system for strategies that trade linear assets, mainly stocks and futures. In this chapter we do not contemplate giving a full and detailed description of a universal backtesting system—such general review may be found in a vast number of books and articles. Instead, we concentrate on specific features which must be considered when option trading strategies are tested.

5.1. Database

Historical data are the cornerstone of any backtesting system. More than 2,000 stocks with options are traded at U.S. exchanges alone. If we assume that each stock has on average about ten actively traded strikes and about ten expiration dates, then we get about 400,000 trading instruments. Such diversity explains special requirements to the organization of the database structure.

5.1.1. Data Vendors

Most brokerage firms providing online access to stock exchange data allow observation of current option quotes. However, granting historical data of sufficient length is not part of the usual brokerage service package. Current quotes for most option contracts can be found at the popular free services of *finance.yahoo.com*,

finance.google.com, and many others. Things get much more complicated with options price history. To create and maintain a historical database, the system developer should apply to special data vendors. Not all of them support a sufficient data length. Since creating backtesting systems for option strategies is a rather complicated task, until recently, the demand for extensive option data was quite low.

Many data vendors offer option price history only for a small range of underlying assets (for example, basic indexes, futures, and selected stocks). Developers of some analytical platforms for option trading (such as the *OptionVue* system) provide access to price history directly from their software products. This might be convenient for the backtesting of simple strategies that can be realized directly within the vendor's platform. However, development of complex strategies based on a customized backtesting system is impossible within the framework of such platforms.

The fairly wide range of underlying assets and a quite considerable length of price history can be found at the Internet resources www.historicaloptiondata.com, www.ivolatility.com, www.livevol.com, www.optionmetrics.com, and www.stricknet.com. In particular, livevol.com offers not only daily close prices but some intra-day prices as well. Apart from price history, IVolatility.com offers a wide range of option-specific indicators and analytical materials.

If the database is purchased as a ready-made product, one should pay special attention to the problem of omission of underlying assets which existed earlier but are no longer traded *(survival bias problem)*. This may be the consequence of company acquisition, bankruptcy, or delisting. A survival bias problem arises if the vendor provides data only for currently existing companies. If "extinct" underlying assets are omitted, strategy backtesting will overlook important market events and backtesting will not be complete and reliable. For example, when there were rumors about a company's possible acquisition, the price of its options could have been very high. If the volatility selling strategy had generated a short position opening signal at that time, considerable price movement (that could have been realized after the event has taken place) would lead to a serious loss and, probably, result in total ruining of the trading account. If the database used for backtesting does not include the ticker of the acquired company, then this event will not be reflected in the backtesting results. Therefore, it is absolutely necessary to use the full set of all underlying assets that have existed during the backtesting period.

5.1.2. Database Structure

The database must contain the minimum information related to a certain time frame (in most cases the frame is one trading day). Regarding the price of any financial instrument, the minimum information set includes open and close prices of the day, as well as daily high and low prices. More detailed databases include intra-day prices recorded with a given frequency, and the most complete bases may contain ticks data, reflecting all deals executed within the given time frame.

Historical prices of options have their own distinctive features. Since options are derivative financial instruments, their pricing is inseparable from the prices of their underlying assets. This leads to a fundamentally important task of *synchronization*—option quotes recorded at a given moment in time must be connected with their underlying asset price recorded strictly at the same moment. This factor must be considered when the structure of the historical database is set.

In view of the wide spreads and due to the low liquidity of options, their bid and ask quotes contain much more information than prices of executed transactions. It is especially important when close prices of a trading day (or another time frame) are determined. In many cases the last option transaction is executed before the trading is closed, while for the underlying asset the last transaction is usually related to the moment of market closing. Since such mistiming is inadmissible, it is preferable to use the last bid and ask option quotes instead of close prices.

Testing most of the option strategies requires the following information entities to be available in the database structure:

- *Underlying assets price history,* including the standard set of open, close, high, and low prices for supported time frames.
- *Options prices history and history of bid/ask quotes.* Each option contract has its own ticker and transactions history. A full-fledged backtesting system requires maintaining a complete structural description linking each option to its underlying asset and to other options related to the same underlying asset.
- *Trading volume* (for both underlying assets and options) and *open interest* (for options).
- *History of dividend payments, splits, and ticker changes* for underlying assets.
- *History of quarterly earnings reports.* The goal of backtesting

is to simulate the past decision-making process without looking into the future. Utilization of the real chronology of quarterly earnings reports (as provided by many data vendors) may cause fundamental errors in the backtesting. In most cases the decision-making process is based on the *expected* date of the event. If the backtesting system is based on the chronology of reports (as realized), it will generate transactions that would never be executed in reality. To avoid such distortions, the *expected* dates of earnings releases should be stored (in addition to the real dates) in the database. These data can be accumulated by automatic daily scanning through financial sites like www.marketwatch.com, www.earnings.com, and www.finance.yahoo.com.

- *History of irregular corporate events,* such as mergers/acquisitions, bankruptcies, court decisions, or announcements of products approval or rejection. These events significantly complicate the backtesting process. While rumors related to such events strongly influence the implied volatility of stock options, information about them becomes reliable only after the event has taken place. However, in the backtesting we should use only assumptions and expectations that prevailed in the market before the date of occurrence of the event. Collecting, storing, and using such information are all extremely difficult tasks. If it is impossible to determine the information that was available to market participants at a given moment in the past (before the event), the transaction should be excluded from the backtesting.
- *Fundamental financial indicators,* like P/E, PEG, ROE, EPS, and other multiples derived from corporate financial statements, which form the basis for fundamental analysis. Using fundamental indicators in option trading is not a common practice yet. However, their application in the option strategies seems to be a promising direction in future development (especially in combination with technical analysis). The data from quarterly reports should be accumulated and linked to estimates by the leading financial analysts (released before the report was published).

Many data vendors provide a whole range of indicators along with price and volume data. This includes different variants of volatility calculations (both historical and implied), volatility indexes, the Greeks, and so forth. Although such information can be included in the database, the trading system developer can use his own models

and algorithms to calculate these characteristics (as the need arises).

5.1.3. Data Access

Simulation of long-term trading requires successive processing of data relating to thousands of trading days. Numerous historical simulations are performed not only in the process of backtesting but also in the course of preliminary strategy optimization and the statistical analysis of its parameters. The data used in these procedures have a very significant volume. Access to such extensive data is fraught with considerable time costs.

The backtesting system works simultaneously with a large number of trading instruments. The instrument may be a separate option or an option combination (a virtually unlimited number of option combinations can be created using separate option contracts). Therefore, apart from simple navigation in time, the system should ensure fast navigation through the option series structure. In particular, the trading algorithm of the tested strategy may require creation of the following dataset on each simulation day:

- All options relating to a certain underlying asset and with a given strike price (or a range of strikes) and expiration date (or several dates within the given time interval)
- All strikes and expiration dates corresponding to a given underlying asset
- All actively traded options (with an average daily trading volume exceeding the predetermined threshold value) relating to a given set of underlying assets
- At-the-money, out-of-the money, or in-the-money options corresponding to a given expiration date
- Many other more complicated sets

The system also needs to be able to determine quickly whether certain statements are true or false—for example, whether the corporate earnings report is expected between the current and the expiration date.

To perform these multidimensional tasks requiring navigation through both time and combination structure, we need to ensure a sufficiently high speed of data access. The adequate data access speed can be secured only by storing all the data required during strategy simulation in the main memory (random access memory). This means that, apart from the main database, intended for accumulation and storage of the extensive historical data, the system should create a temporary auxiliary database that provides prompt access to the

specified data at each step of simulation. We will refer to such an object as “the operative history.”

The operative history represents an object that contains only relevant information that is required for the backtesting of a certain strategy. It is loaded at the beginning of the backtesting process and significantly diminishes the problem of access rapidity. For example, if the strategy uses only indexes or ETFs as underlying assets, there is no need to utilize the datasets related to expected quarterly earnings reports (these data are contained in the main database, but not in the operative history). The strategy can also be limited to using just the several nearest expiration dates and strikes situated within quite a narrow range around the current underlying assets prices. Even despite such limitations, placing the operative history in random access memory may require application of the data compression algorithms.

5.1.4. Recurrent Calculations

Backtesting of strategies with a focus on trading stocks, futures, or other linear assets uses many technical and fundamental indicators. In the backtesting of option strategies, other specific characteristics are calculated in addition to the traditional indicators—valuation criteria, implied and historical volatility, and Greeks relating to the whole range of instruments. The extensive volume of these data does not always allow calculation of the necessary indicators beforehand and then storing them in the database. To avoid overloading the database with superfluous data (which might only occasionally be needed), these indicators are calculated as the need arises. However, such an alternative approach (calculating instead of storing indicator values) leads to another problem: excessively burdensome time and computational resource costs. To solve this problem, special techniques of calculations acceleration are required. One of these technologies is recurrent calculations. In particular, recurrent calculations can effectively be applied for historical volatility.

Historical volatility is frequently used to calculate criteria that require integration of the payoff function with respect to the probability density function of the underlying asset price. The typical criteria of such kind are expected profit and profit probability calculated on the basis of lognormal distribution. In the course of strategy backtesting, these criteria are calculated for each moment in time t. Hence, values of historical volatility $HV(t)$ have to be calculated for each moment in time. It is preferable to calculate all volatility values for all underlying assets and all moments in time before the simulation begins. This can be done by applying the

recurrent calculations. Since the historical volatility corresponding to certain time t is

$$HV(t) = \sqrt{\frac{r^2(t-N+1) + r^2(t-N+2) \ldots + r^2(t)}{N}}, \quad r(t) = \ln \frac{C(t)}{C(t-1)},$$

where $C(t)$ is the closing price of the t-th day and N is the length of historical period used to calculate volatility, historical volatility corresponding to the next day is

$$HV(t+1) = \sqrt{\frac{HV^2(t) \cdot N - r^2(t-N+1) + r^2(t+1)}{N}}.$$

Obviously, the complexity of calculations is almost independent of N, given a sufficient length of the historical database. Using recurrent calculations enables us to calculate the function value for each following day on the basis of the previous value of this function. This significantly reduces the number of calculations done. A similar approach can be applied for any volatility formula with reversible function F of the following type:

$$HV(t) = F(\sum_{i=1}^{N} f(r(t-i)).$$

5.1.5. Checking Data Reliability and Validity

Data obtained even from the most reliable sources may contain errors. Besides primitive errors originating from software bugs (for example, a decimal point is replaced by a comma, a ticker is misspelled, or split information is omitted), there may be systematic errors related to false or missing quotes. Serious problem arises due to desynchronization of the underlying asset price and prices of its options. Trading simulated on the basis of erroneous or desynchronized data is fraught with severe distortions of results.

The problem of erroneous data should be solved by way of their detection and further correction or filtration. The detection procedures should be applied to all new prices and bid/ask quotes entering the database. There are two main approaches to detecting erroneous data: tests for availability of arbitrage situations and analysis of implied volatility surfaces.

Tests for arbitrage situations are based on the theory of options pricing and the concept of option prices parity. For European options with the same expiration date, put/call parity is expressed by the following equality (provided that no dividends are expected before the expiration date):

$$C + \frac{X}{(1+r)^t} = P + S \,,$$

C and P are prices of the call and put options, respectively, X is the strike of these options, S is the current underlying asset price, r is the risk-free rate, and t is the time left until options expiration. The parity notion is based on the principle that arbitrage situations cannot exist in an efficient market. Accordingly, there is no financial instrument or set of instruments that would generate return in excess of the risk-free rate without accepting any financial risk.

If parity is violated, this indicates that options prices are false or there is information "priced in" by the market but disregarded by the testing algorithm (for example, dividend payments are expected). The error may be in the price of one of the two options or in both of them. In the latter case we need to find out which of the prices is erroneous. Each of these two options should be examined separately, paired with another option that successfully passed the parity test. This other option will inevitably have a different strike price or a different expiration date. Once the option with the erroneous price has been detected, it should be excluded from the database or its price should be corrected by solving the parity equation (the price of the erroneous option is set to be the unknown and the correct price is fixed as the constant).

The significant drawback of this method is that it is applicable only to European options. For American options the parity is expressed by the following inequality:

$$S - X \le C - P \le S - \frac{X}{(1+r)^t} \,.$$

As it follows from the formula, there is no exact parity value for American options. The testing algorithm can only determine whether the difference between put and call prices lies within the permissible range. Although going beyond the limits of this range indicates parity violation, being within the range does not necessarily mean that the possibility of price distortion is ruled out. The price of one of the options (or both) may be erroneous, but not enough to get beyond the range. However, prices of European options can also be distorted without violating the parity. For example, if the put price is decreased by some amount, while the call price is increased by the same amount, the parity will still hold despite the fact that the prices of both options are incorrect.

Another method of detecting erroneous option prices is based on *implied volatility estimations.* Values of implied volatility have to be

calculated for all options corresponding to a given underlying asset. Due to many different reasons, these values will not coincide for options with different strikes and expiration dates. At the same time, their divergences are not random. In particular, the relationship between implied volatility and the strike price frequently has the shape of a smile (it is called “volatility smile”)—the lowest volatility value corresponds to the strike that is nearest to the current underlying asset price; as the strike moves away in both directions (toward both the in-the-money and the out-of-the-money directions), the implied volatility increases. There may be other forms of the relationship. However, whatever the shape of the relationship, it represents a more or less smooth curve (or, in some cases, a straight line). If one of the volatility values falls out of the general trend and is situated far from its expected place on the curve, the price of the option corresponding to this strike is probably distorted. A similar relationship can be established between implied volatility and the expiration date. Again, price distortions can be detected by analyzing the smoothness of such a relationship.

In many cases it is preferable to analyze the volatility surface. It is constructed by establishing the relationship between implied volatility, strike, and expiration date. Under normal conditions this relationship has a shape of a relatively smooth bending surface. Sharp peaks and lows on this surface point to possibly incorrect values of implied volatility, which means that the option price might be distorted. The main problem with this method is that detection of anomalous prices is based on visual analysis of the volatility surface chart. This task becomes absolutely unfeasible when the database is regularly replenished with thousands of new tickers. This problem can be solved by organizing an automated search for the outlying points, though creating such algorithms is not a trivial task.

Within the framework of a two-dimensional linear relationship (between volatility and strike or between volatility and expiration date), this problem is solved easily by linear approximation. Anomalous points are found by comparing their residuals (obtained from the regression model). If the residual of some point differs from the average residual by more than a threshold value, this may signal the distortion of the corresponding option price. In those cases when the relationship is nonlinear (with a different form of volatility smiles), more complicated curve fitting techniques should be applied.

The problem becomes even more pronounced when we move on from the two-dimensional system to the volatility surfaces. In these cases more complicated mathematical models must be applied. The main requirements to these models include high precision in data

approximation and availability of the analytical formula describing the volatility surface. The formula is required to calculate the correct price in case the distorted price(s) is (are) detected. The correct price is estimated by eliminating the outlier from the model and recalculating the surface formula. Then the new value of implied volatility is calculated for the outlier by applying the new formula. After the volatility value has been corrected, we can use it to correct the erroneous option price (this is done by applying one of the option pricing models).

5.2. Position Opening and Closing Signals

The algorithm of the trading strategy generates formalized decisions on opening and closing trading positions. The modeling of this decision-making process requires calculation of the values of special *functionals* intended to evaluate profitability and the risk of potential trading variants. Further processing of functional values generates the position opening and closing signals, which are transformed later into trading orders.

5.2.1. Signals Generation Principles

In options trading, underlying assets and separate option contracts form a universe of financial instruments that are available for implementing different trading strategies. At the same time, option combinations may also be considered as trading instruments. In fact, we are not obliged to put any constraints on the complexity of the trading instruments. For example, any set of option contracts relating to the same underlying asset can be treated as a separate trading instrument.

Let $\Omega = \{K_1, \ldots, K_N\}$ be a set consisting of N trading instruments. At any moment in time t (that is, at each step of simulation), the strategy disposes of a certain information set $I(t)$ that is currently available. This information is very diverse: price history of relevant trading instruments, formalized fundamental information, technical indicators, profits/losses of positions that have been opened earlier, different probabilistic scenarios, and so forth. The only requirement to $I(t)$ is that data from future moments in time $t+1$, $t+2$,... should never be included in any form into this information set. For each trading instrument the strategy algorithm calculates special functionals $\Phi(K_j, I(t))$, where K_j is the j-th trading instrument. Interpretation of these functionals is the element of the trading strategy.

The functional may assume a logical value. For example, logical values *false* and *true* may have the following meaning: $\Phi(K_j, I(t)) =$

true means that the strategy has generated a signal to open a trading position for the instrument K_j. This may be either a buy trade, if the functional $\Phi(K_j, I(t))$ is responsible for generating buy signals, or a sell trade if this functional generates sell signals. $\Phi(K_j, I(t)) = false$ means that the strategy has generated no signal. The value of functional Φ may be real numbers. In this case the functional value reflects not only the presence or absence of a signal, but its relative strength as well.

Suppose that functional Φ evaluates the profitability of potential trades for N instruments. Assume further that at step T the vector of functional values $\overline{\Phi} = \{\Phi(K_j, I(t)), j = 1,\ldots, N\}$ is obtained. The next step of the algorithm is the interpretation of these values and generation of the vector of signals that will be used at the next step for generation of trading orders. Let us consider two methods of transforming the vector of functional values $\overline{\Phi}$ into the buy and sell signals.

The first method is based on using threshold values. Suppose the long position is entered, if the value of the functional, which is based on the profitability criterion (expressed as a percentage of investment capital), exceeds the threshold of 3% ($\theta_{buy} = 3$). If this criterion forecasts losses exceeding 3% ($\theta_{sell} = -3$), the short position is entered. According to the threshold-based method, the opening signals are generated in the following way:

$$\textit{buy instrument } K_j \textit{ if } \Phi(K_j) > \theta_{buy},$$

$$\textit{sell instrument } K_j \textit{ if } \Phi(K_j) < \theta_{sell}.$$

If these inequalities do not hold (that is, $\theta_{buy} > \Phi(K_j) > \theta_{sell}$), trades with instrument K_j are not initiated.

The second method is based on ranking trading instruments according to the order of elements of the vector $\overline{\Phi}$: $\overrightarrow{\Phi} = (\Phi(K_{j_1}), \Phi(K_{j_2}),\ldots, \Phi(K_{j_N}))$, where $\Phi(K_{ji}) \geq \Phi(K_{ji+1})$. Such ranking enables us to pick out a certain number a of superior and b of inferior instruments: $Top = \{K_{j1}, K_{j2},\ldots, K_{ja}\}$ and $Bottom = \{K_{jN-b+1}, K_{jN-b+2},\ldots, K_{jN}\}$. The position opening signals are generated in the following way:

buy all instruments belonging to the set Top,

sell all instruments belonging to the set Bottom.

These procedures were simplified in order to demonstrate the main approaches to generation of trading signals. In reality, more complicated algorithms are used, though their core principles are very

similar to our description.

5.2.2. Development and Evaluation of Functionals

Different valuation criteria described by Izraylevich and Tsudikman (2010) can be used as functionals for generation of position opening signals. These criteria express quantitatively the attractiveness of option combinations (when they are used as trading instruments). Regardless of the peculiarities of any particular functional, its application in the backtesting system is reduced to the evaluation of potential profitability and risk of financial instruments.

Continuous search for efficient functionals is one of the main tasks performed by the strategy developer. A range of statistical research is used as the basis for the construction of new functionals and elaboration of existing functionals. To carry out such research, a multitude of trades are modeled during a certain historical period by applying the functional under investigation. Statistical properties of this set of simulated trades—average profit, variance, distributions, and correlations—are analyzed. On the basis of this analysis, effective functionals can be selected for further examination.

However, having good statistics is a necessary, but insufficient, condition to judge the functional as possessing high forecasting qualities. The order of profitable and unprofitable trades is also of great importance for a strategy to be successful (statistical research disregards this order). Thus, a long series of unprofitable trades may be practically unacceptable because of high capital drawdown or an excessive length of the unprofitable period. Furthermore, the strategy that has performed satisfactorily in the past, but has recently turned into an unprofitable one, may also show good statistics. Therefore, statistical research may and should be used to create functions, while historical simulations and backtesting should be used to optimize their parameters and to evaluate their effectiveness.

5.2.3. Filtration of Signals

Depending on the strategy algorithm, some trading signals should be eliminated without transforming them into position opening orders. This might be necessary either because of the upcoming corporate events or due to the fact that the values of some indicators do not satisfy specific requirements of the strategy. These indicators are intended to filter out the most risky option combinations.

Let us consider the example of the risk indicator which may be efficiently applied for the filtration of unacceptable trading signals. The asymmetry coefficient (described in Chapter 3, “Risk Management”) evaluates the extent of asymmetry of the option

combination relative to the current underlying asset price. If expected profit, calculated on the basis of empirical distribution, is used as the functional evaluating the attractiveness of option combinations, many asymmetrical combinations may turn out to be very attractive (they may have high expected profit values). This happens when empirical distribution also has an asymmetrical shape. Since expected profit is calculated by integrating the combination payoff function with respect to the probability density function of empirical distribution, the integral value is high when both functions are asymmetrical and their modes are skewed to the same side. Nevertheless, such combinations are inappropriate for strategies based on short option selling. Due to their asymmetry, the premium received from options selling consists mainly of intrinsic value, while the time value is quite low. Hence, the profit potential of such combinations is limited. Application of the "asymmetry coefficient" indicator allows filtering position opening signals that relate to such combinations.

Filtration that is required because of the forthcoming corporate events may be executed in two ways. The *direct filtration method* is applied when information concerning the expected events is accumulated and maintained in the database. The underlying assets, for which significant events are expected, may be temporarily excluded from the database (trading signals for such instruments will not be generated). Such events include quarterly earnings reports, since information about the event date is usually available and more or less reliable.

The *indirect filtration method* is applied when information about forthcoming events is not found in the database. In such cases we use special filters that eliminate signals generated for instruments for which an unusual event is pending. One such filter is the ratio of implied to historical volatility *IV / HV*. Historical volatility *HV* of an underlying asset characterizes the variability of its price during the period preceding the current moment. Implied volatility *IV* expresses the price variability of the same underlying asset that is expected by the market in the future. When no important event is expected, these two volatilities have close values (the ratio is close to 1). If this ratio deviates significantly from 1, this may indicate that an important event with an uncertain outcome is pending.

If implied volatility significantly exceeds historical volatility ($IV / HV > 1$), we must conclude that the underlying asset price was less volatile relative to fluctuations that the market expects in the future. For the backtesting system this suggests that some important event that might significantly influence the price is forthcoming. Among the typical examples of such events are court decisions, news on mergers

and acquisitions, and approval of new products. Unless the strategy is explicitly based on such kinds of events, these underlying assets should be filtered out before the backtesting starts.

If historical volatility significantly exceeds implied volatility ($IV / HV < 1$), we must conclude that some important event has already taken place and that the market does not expect any further price movements for this underlying asset. In most cases combinations corresponding to such underlying assets do not match the basic parameters of the tested strategy and should be excluded from the backtesting system.

5.3. Modeling of Order Execution

After opening and closing signals have been generated, they are converted into virtual trading orders. This section is dedicated to simulation of order executions. Low liquidity of options may be a serious obstacle to successful execution of orders generated by the trading strategy. In the case of limit orders, this will lead to partial execution. In the case of market orders, low liquidity may cause considerable price slippage resulting in a worse (than expected by the strategy algorithm) execution price.

Effective backtesting is possible only when simulated trades do not differ from executions that are attainable in real trading. Although some deviations in execution prices and volumes are inevitable, the system developer should strive to get them as close as possible to the reality. To minimize the discrepancies between simulated and real trades, the backtesting system should include algorithms for modeling partial execution of trading orders and execution using prices that differ from those in the historical database. Besides, the execution price should be adjusted to account for fees and commissions that are charged for the execution of trading orders.

5.3.1. Volume Modeling

It is highly desirable for the historical database to contain bid and ask volumes. However, the execution volume of a limit order can hardly be assumed to match the volume of the corresponding quote. The quote volume appearing in the database may aggregate volumes from different exchanges. For capturing this volume the order has to be split by the broker into several parts and sent to different exchanges. Besides, the option market is based on a market-making principle, which allows the dealer to continuously change bid and ask volumes depending on many factors: the market situation, his own holdings, and even the current stream of trading orders he is entrusted with. Therefore, quotes volumes contained in the database may be

used only as one of the input components required for estimating the execution volume of limit orders.

Other information that can be useful for forecasting the execution volume is the daily volume of executed trades and open interest. However, these data should also be treated with caution. Trading volumes relating to specific option contracts can be very variable and sporadic. Large trades may occur episodically and not reflect the real market depth. The same is true for the open interest, which may be inflated by a few large trades that have been executed in the past and should not be expected to persist in the future.

Simulation of partial order execution can be realized using different estimation methods. However, the outcome of all methods is reduced to determining the value of one parameter: the percentage of order execution. This parameter can be viewed as a function with the following arguments: the current bid and ask volumes, the average daily trade volume (calculated using a certain number of previous trading days), the open interest, bid/ask spread, the distance between the strike price and the current underlying asset price, and the number of days left until option expiration.

The specific formula employed for calculating such a function and history length that is used to average daily volumes may vary. For example, the developer can account for the growth of option trading volumes that usually occurs as the expiration date approaches or as the underlying asset price approaches the strike of the option.

Alternatively, the developer of the backtesting system can simulate volumes of order executions as a random variable changing according to a specific distribution law. The parameters of such distribution should be derived from the same characteristics (bid and ask volume, average daily volume, open interest, spread, etc.).

When the execution volume is modeled, it should be taken into account that the order generated by the backtesting system must be a multiple of the standard lot amount. While theoretically a single option can be considered as one separate security, in reality options are traded in standard lots, the amounts of which are determined by the exchanges. Usually, one lot of options on U.S. stocks is 100, on S&P 500 E-Mini futures it is 50, and on VIX futures it is 1,000. These multipliers are a necessary element of the option database. The volumes of simulated trades must always be multiples of lots.

5.3.2. Price Modeling

Under the low liquidity conditions the limit order may fail to be executed in full (the executed volume is lower than the order volume),

though its price cannot be worse than the filed limit. The opposite situation is true in the case of the market order: The execution volume usually matches the order, while the execution price may be worse than the market price that existed at the moment when the opening signal was generated. This phenomenon is called “slippage” or “market impact.”

To simulate the execution price, we need to introduce a parameter expressing the slippage magnitude into the backtesting system. By analogy with volume simulation of the limit order, this parameter can de presented as a function with the following arguments: the bid and ask volumes, the average daily trade volume, the open interest, the bid/ask spread, the distance between the strike price and the current underlying asset price, and the number of days left until option expiration. In most cases slippage is lower when the average daily volume is higher and the spread is lower. The closer the strike price is to the current underlying asset price and the fewer the days that are left until the expiration date, the higher the options liquidity will be (consequently, the slippage is lower).

When the developer constructs strategies intended for trading highly liquid options with at-the-money strikes and close expiration dates, the impact of slippage can be neglected. When such strategies are backtested, the volume of positions that can be opened during a given period of time for each contract should not exceed a certain share of average daily volume. Specific values of the time period and the percentage of average daily volume are the parameters (they must be set by the system developer depending on his execution facilities).

If slippage is assumed to be zero, the worst spread side is usually used as the execution price for market orders (the ask price for buying, the bid price for selling). However, we can assume further that, in real trading, orders can be executed at better prices than the worst spread side. The execution system can be automatic (based on special algorithms) or can be assigned to a trader. Various intermediate alternatives—when orders are executed by a human with the help of different automated algorithms—are also possible. Depending on trader qualification and the effectiveness of applied algorithms, the execution system may be able to execute orders at the prices that are inside the spread.

This possibility can be taken into account by introducing parameter μ (with values ranging from 0 to 1) into the backtesting system. When $\mu = 0$, the worst execution prices are used—the ask price for buying, and the bid price for selling. When $\mu = 1$, the best execution prices are used—bid for buying and ask for selling. In general,

$$sell\ price = \mu \cdot Ask + (1 - \mu) \cdot Bid,$$
$$buy\ price = \mu \cdot Bid + (1 - \mu) \cdot Ask.$$

Since option spreads are usually wide, the influence of parameter μ on the performance of the option-based strategies may be substantial. In some cases its impact may be so significant that even a slight change in μ can turn an unprofitable strategy into a profitable one (and vice versa)! Thus, it is highly important to select a reasonable value for this parameter. The best solution would be to collect real empirical data of trading order executions. With these data in hand, the developer can investigate the relationship between the real execution prices and prices that were used by the backtesting system to generate opening and closing signals. On the basis of this relationship, it would be possible to estimate the appropriate value of the μ parameter (this value is specific for each particular execution system).

5.3.3. Commissions

Every transaction simulated by the backtesting system is recorded on the virtual brokerage account and is further used to evaluate strategy performance. Apart from execution volume and price, the transaction is characterized by the commission paid to the broker. The terms of commissions depend on the type of securities, exchange fees, volumes and frequency of trading, and individual broker terms. There are several approaches to calculating commissions. Usually a client of a brokerage firm can choose the most appropriate alternative, depending on the nature of his trading operations. The following three alternatives are the most common:

1. Commissions are proportional to the number of securities purchased or sold.

2. Commissions are proportional to the transaction amount.

3. A flat fee is set for a certain period (for example, a month) and does not depend on trading turnover.

Commission rates usually have fixed minimum and (sometimes) maximum limits and may vary depending on the type of securities. In any case, the formula for calculating the exact value of trading commissions must be integrated into the backtesting algorithm (it may vary depending on the strategy under evaluation).

In option trading commissions may be very high and in some cases even absorb up to 50% of strategy profit. There are several reasons for that. Firstly, in contrast to positions in stocks or futures, the option position often consists of a whole set of instruments (a combination of different option contracts), each of which is a separate security

requiring execution of a separate trade. Since commissions may have minimum limits for each operation, establishing a single option position may generate more commissions as compared to positions in other financial instruments. Secondly, option positions in many cases give birth to positions in underlying assets, which require payment of additional commissions to close them. Thus, the backtesting system should provide for analysis of the sensitivity of strategy performance relative to changes in the commission rate.

5.4. Backtesting Framework

In developing a backtesting system the developer struggles to solve two main problems: the thorough evaluation of the strategy performance (this is examined in section 5.5) and evaluation of the likelihood that the performance shown using historical data will hold in future real trading. In fact, this likelihood is impossible to evaluate in terms of probability. It would be more appropriate to describe the second problem in the following way: to maximize the probability that historical strategy performance will not deteriorate significantly in real trading.

To create a sound backtesting system that would ensure strategy continuity and integrity, the developer has to construct a solid framework that is based on the following interrelated principles:

- Reasonable separation of the historical database into a period when the strategy is optimized (the period of *in-sample* optimization) and a period when the strategy is tested (the period of *out-of-sample* testing)
- The possibility of recurring optimizations—the optimization procedure should be repeated continually as the strategy advances through price history *(adaptive optimization)*
- Adequate handling of potential overfitting challenges
- Special methods for testing the robustness of the backtesting system

5.4.1. In-Sample Optimization and Out-of-Sample Testing

The historical period covered by the available database is divided into two parts: the strategy optimization period τ_s (in-sample) and the strategy testing period τ_o (out-of-sample). The positional relationship of the in-sample and the out-of-sample periods can be different. Period τ_s may precede period τ_o or vice versa. The continuity of periods is also not obligatory.

Period τ_s is used for creating the strategy. During this period, the basic idea is tested, the set of optimized and fixed parameters is

determined, and optimal parameter values are found. Different optimization techniques are used here. After the final variant of the strategy has been selected, its efficiency is verified by running at the interval τ_o, which was not used in the process of strategy creation. The same indicators that have been used when the strategy was developed are measured here (these indicators reflect various forms of profitability, risk, stability, and other properties of the tested strategy). If the values of these indicators show no significant deterioration (that is, they remain within acceptable limits), we assume that the strategy is stable and can maintain its effectiveness in future trading.

The main problem associated with historical time series is their nonstationarity. Market behavior differs depending on the prevailing sentiment of market participants, and the economic and political situation. Phases that are commonly encountered in the stock and futures markets can be classified into three categories: upward trends, downward trends, and periods of relative price stagnation (lacking any pronounced trend). In the options market we should also consider phases of high and low volatility. The longer the period of strategy optimization, the higher the probability that this period will cover different market phases.

Any given strategy may show different efficiency during different market phases. Even if we do not strive for the creation of universal strategies performing well under any market conditions, it is still important to ensure that the profitable strategy will not turn into a frustrated one when the market phase changes. By prolonging period τ_s, we increase the probability of capturing diverse market phases, and, thus, reduce the sensitivity of the strategy to the changing market environment. Accordingly, optimization performed using a longer time period is more *reliable*. On the other hand, the shorter the optimization period, the better the strategy is customized to the conditions prevailing in the market in recent time. Other things being equal, such optimization is more *stable*. Thus, the decision about the length of the historical period used for strategy optimization is a compromise between reliability and stability.

We suggest a guideline that can help find a trade-off between reliability and stability for option trading strategies. It is based on the weighted (by position volumes) average time interval between the moment of portfolio creation and that of options expiration. In general, the longer the period until expiration, the longer the period of optimization should be. For example, if the strategy opens positions that consist mainly of options with 2 to 20 days left until expiration, the optimization period may be much shorter than in the case of the

strategy that is based on trading longer-term options (for example, LEAPS).

The length of the testing period should not be less than the period left until options expiration. We can propose the following rule-of-thumb: τ_s should include at least ten full non-overlapping expiration cycles. For example, if the weighted-average time interval between the moment of portfolio creation and that of options expiration is 15 working days, the minimum τ_s length should not be less than 150 days. Such estimations should be adjusted depending on the average number of positions that are opened during one expiration cycle. The higher the number of trading operations executed by the strategy, the higher the credibility of performance statistics and the shorter period τ_s may be acceptable.

5.4.2. Adaptive Optimization

The algorithm of the trading strategy may provide for the *adaptive optimization procedure*. This is realized as a moving optimization window. To formally describe the adaptive optimization, let us consider the description of strategy $S(P)$, where P is the vector of strategy parameters. Let $\tau(T) = [T - \Delta t + 1, T]$ be the historical interval ending at some moment T and having the length of Δt days. When point T moves from the past to the future, the interval $\tau(T)$ also moves after T. Assume that optimization algorithm A generates the vector of parameters $P^*(T) = A(\tau(T))$. Let l be the distance between optimization moments, and $T0$ be the initial moment in time. Then the trading algorithm with adaptive optimization is the following. At moments $T_n = T0 + nl$ (where n is the serial number of the adaptation step with values $n = 0,1,2,\ldots, n_{last}$, the optimization algorithm A is activated to generate new parameter values $P^*(T_n) = A(\tau_0(T_n))$. The strategy $S(P^*(T_n))$ trades during the next after time interval $[T_n + 1, T_{n+1}]$, after which a new step of adaptation and trading is made.

As a result of introducing the adaptive optimization procedure, the original strategy $S(P)$, which initially had the set of parameters P, turns into a complex strategy that can be described by the sequence $\{S(P^*(T0)), S(P^*(T_1)),\ldots,S(P^*(T_n)),\ldots\}$. This complex strategy has two additional parameters, l and Δt, that may also be optimized in order to find their best values. Performance of the strategy providing for periodical reoptimization may be evaluated using standard effectiveness indicators (see section 5.5). Such evaluation enables us to judge the usefulness of adaptive optimization in each specific case.

In many cases adaptive optimization allows creating strategies that are more robust (that is, less sensitive) to changing market phases.

However, it should be clearly noted that adaptive optimization does not eliminate the overfitting problem—it is still just an optimization, although with a more complicated structure. Moreover, using adaptations in the backtesting system inevitably leads to increasing the number of optimized parameters, which might increase the overfitting risk (see section 5.4.3). Nevertheless, our experience suggests that when the strategy is reoptimized, its future robustness is higher than it would be if it had been optimized at an unchanging historical period (provided that it is tested during a sufficiently long historical interval and generates a sufficient number of position opening signals).

5.4.3. Overfitting Problem

The high effectiveness demonstrated by many strategies (in simulations that are based on past historical data) is frequently attributed merely to the excessive optimization of their parameters. Usually, performance of the same strategies deteriorates significantly when they are tested using the out-of-sample price series. This phenomenon is known as the overfitting or curve-fitting problem. Since no trading algorithm is possible without using (either explicitly or implicitly) some parameters, the overfitting risk is inevitable even for relatively simple strategies. For more complex strategies this risk increases manifoldly.

In general, we can claim that the probability of overfitting is directly related to the number of degrees of freedom in the backtesting system. In the backtesting of strategies oriented toward trading stocks or futures, the number of degrees of freedom is equal to the number of optimized parameters. For option strategies the number of degrees of freedom is higher because, apart from the parameters relating directly to the trading algorithm, there are some additional parameters related to the construction of option combinations. Option-specific parameters include, for example, the range of allowable strike prices, limits imposed on the time interval between the moment of position opening and the expiration date, the length of historical horizon that is used to calculate volatility, and many other parameters. Thus, for option-oriented strategies the overfitting problem is even more relevant than for strategies used in the trading of plain assets.

Unfortunately, there is no way to fully eliminate the overfitting risk. However, this problem can be substantially diminished, if the following general principles are observed.

Walk-forward analysis is the main method of struggling with overfitting. This method is based on dividing historical time series into periods of optimization (in-sample) and testing (out-of-sample) (see

section 5.4.1.). The main idea is that the evaluation of strategy performance does not involve data that have been used in strategy creation and optimization. However, it is necessary to note that satisfactory performance, recorded during out-of-sample testing, does not necessarily prove that the strategy is not vulnerable to the overfitting risk. Although this statement seems to be contradictory at first sight, this can be explained with the following reasoning. Suppose that the developer sets two requirements for the backtesting system:

1. Values for strategy parameters that cause the objective function (or functions) to achieve satisfactory value(s) during the in-sample period should be found.

2. These values should not deteriorate during the out-of-sample period by more than a given amount.

It was shown earlier (see Chapter 2, "Optimization") that quite a few parameter combinations with acceptable objective function values (so-called optimal zones) may be identified in the course of optimization (using the in-sample period). Consequently, it may well turn out that at least one variant from this multitude will give (solely by chance) satisfactory results during out-of-sample testing as well. This means that good results obtained via walk-forward analysis may be merely the result of fitting (overoptimization) and will most likely not hold in future real trading.

When the strategy is tested using walk-forward analysis, the number of trades should be thoroughly controlled. The appropriate relationship between the number of executed trades and the number of strategy parameters must be established. In general, the higher the number of parameters used in optimization, the more trades should be executed. Adhering to this principle ensures minimization of the overfitting risk.

If the strategy is unstable (small changes in parameter values lead to significant deterioration of strategy performance), it is most likely overfitted (different aspects of stability, which is also called "robustness," were discussed in Chapter 2). Two parallel lines of research can be conducted to ensure that the strategy is sufficiently robust relative to the optimal values of its parameters. The first line of research consists in analyzing changes in strategy performance in the neighborhood of the optimal parameter values. The second one is based on investigating the strategy behavior at slightly modified historical data (obtained by applying small distortions to original data). The robustness of the strategy with respect to small changes in parameter values and time series indicates lower overfitting risk.

5.5. Evaluation of Performance

Evaluation of strategy performance is a multicriteria task. There is no single characteristic (the terms “criterion,” “characteristic,” and “indicator” will be used interchangeably) that would allow thorough evaluation of strategy effectiveness and its comparison with other strategies. Instead, there is a whole set of universal indicators that are used in the backtesting of different trading strategies regardless of the kind of financial instruments employed. Although these indicators are widely known, their calculation and application for option strategies have a number of specific features. Apart from universal indicators, there are also some specific indicators that were developed exclusively for the evaluation of option strategies.

5.5.1. Single Event and Unit of Time Frame

To discuss the indicators that are used in the evaluation of strategy performance, we need to give a strict definition of a single event and a unit of time frame.

A single event can be defined as a single trade relating to a specific option contract or as a set of trades relating to the option contract and executed within one day. Besides, a single event may be defined as a set of trades executed within one day to create an option combination. When plain financial instruments (stocks and futures) are traded, each separate trade is usually treated as a single event. In option trading the choice of the single event depends on the peculiarities of the strategy. In most option strategies a set of trades relating to the same combination or the same underlying asset are considered as a single event.

The important factor in determining *the unit of time frame* is the necessity to evaluate the outcome of option trades at the moment of expiration. Option strategies are based on selection of trading variants using special criteria that evaluate option combinations at the moment of position opening. Accurate verification of these evaluations is possible on the expiration date only. The traditional expiration moment of most options traded at CBOE (the main options exchange) is the third Friday of each month. Thus, performance evaluation of many strategies that trade stock options has a natural monthly periodicity resulting from the standard expiration schedule. For such strategies the unit of time frame is one month.

For strategies dealing with long-term options, the expiration moment does not play such a key role. In these cases other reference points (suitable for interim portfolio evaluations) become more important. Recently, options with weekly expirations have been

introduced by exchanges. For strategies employed in trading weekly options, the selection of the unit of time frame of one week seems to be the most appropriate solution.

5.5.2. Review of Strategy Performance Indicators

Strategy performance is measured at the time interval $\tau = [T_0, T_N]$, which is divided into a sequence of interim time moments $\{T_0, T_1, \ldots, T_N\}$. For the sake of simplicity, we will measure time in days (though any unit of time frame may be used for performance evaluation). The length of interval τ is $\Delta T = T_N - T_0$ days, or

$$\Delta T = \frac{1}{A}(T_N - T_0)$$

years, where $A = 365.25$ calendar days. Let $\{E_0, E_1, \ldots, E_N\}$ be the sequence of capital values determined at each of the points in time $\{T_0, T_1, \ldots, T_N\}$. The capital value is usually estimated as the liquidation value of all positions (calculated using close prices of the corresponding day) plus free cash.

Characteristics of Return

We will consider two types of return corresponding to two different approaches to capital management (here we refer to the first level of the capital management system; see Chapter 4, "Capital Allocation and Portfolio Construction"). The first approach requires a constant amount of capital to be invested at each period of time (regardless of the strategy performance during previous periods). The second approach is based on the reinvestment principle.

The first approach is applicable for analyzing the performance of a series of single-type portfolios. It may be appropriate for the strategy that creates one separate portfolio for each expiration date. The same amount of capital E_0 is used for each unit of time frame. Performance of the strategy with constant investment amounts is estimated by calculating *the linear annual return:*

$$r^l = \frac{1}{\Delta T}\frac{E_N - E_0}{E_0}.$$

This characteristic represents the annualized arithmetical mean return.

The second approach is more appropriate for comparing strategy performance with a certain reference return, such as a continuously accrued interest rate (for example, a risk-free rate) or with the performance of some benchmark index (for example, S&P 500). Performance of a strategy that is based on reinvestments of profits generated during previous periods is estimated by calculating *the*

exponential annual return:

$$r^e = \frac{1}{\Delta T} \ln\left(\frac{E_N}{E_0}\right) .$$

This characteristic represents an annualized geometrical mean return.

By linking the moments when capital measurements are taken with option expiration dates, we obtain a series of monthly profits and losses. Let *N* be the number of months in the period of strategy backtesting. According to the first (linear) approach to capital management, the invested capital always equals *E0* and the capital measured at the end of the *i*-th month is *Ei*. Then the profit of the *i*-th month is $p_i^l = E_i - E_0$, the average monthly profit is $p^l = \frac{E_N - E_0}{N}$, the average return is $r^l = \frac{p^l}{E_0}$, and the return of the *i*-th month is $r_i^l = \frac{E_i - E_0}{E_0}$. According to the second (exponential) approach to capital management, the invested capital of each month is equal to the ending capital of the preceding month. There is little sense in calculating the average monthly return, since different amounts are invested in each month. The return of the *i*-th month is $r_i^e = \frac{E_i - E_{i-1}}{E_{i-1}}$, and the average return is the geometrical mean return $r^e = \sqrt[N]{\prod_{i=1}^{N} r_i^e}$.

Sets of $\{p_1^l, p_2^l, \ldots, p_N^l\}$, $\{r_1^l, r_2^l, \ldots, r_N^l\}$ and $\{r_1^e, r_2^e, \ldots, r_N^e\}$ are used to calculate different statistics that are very useful for the evaluation of strategy performance. The maximum monthly profit $profit_{\max}^l = \max_i\{p_i^l : p_i^l > 0, i = 1,2,\ldots,N\}$, maximum linear monthly return $return_{\max}^l = \max_i\{r_i^l : r_i^l > 0, i = 1,2,\ldots,N\}$, and maximum exponential monthly return $return_{\max}^e = \max_i\{r_i^e : r_i^e > 0, i = 1,2,\ldots,N\}$ characterize the most successful month in the time series.

Maximum monthly absolute and relative losses are also very important characteristics. If their values turn out to exceed a certain threshold level, the strategy might have to be rejected. Even a highly

profitable strategy with a single unprofitable month can be rejected, if this loss is unacceptably high. Calculation of the absolute maximum monthly loss $(loss^{l}_{\max} = -\min_{i}\{p^{l}_{i} : p^{l}_{i} < 0, i = 1,2,...,N\})$ makes sense only in the linear case. The relative value for the linear case is $relativeloss^{l}_{\max} = -\min_{i}\{r^{l}_{i} : r^{l}_{i} < 0, i = 1,2,...,N\}$; for the exponential case it is $relativeloss^{e}_{\max} = -\min_{i}\{r^{e}_{i} : r^{e}_{i} < 0, i = 1,2,...,N\}$.

Other indicators may also be used to characterize the strategy return: the number of profitable months, the number of unprofitable months, the average profit of profitable months, the average loss of unprofitable months, the maximum number of successive profitable months, the maximum number of successive unprofitable months, and so on.

Maximum Drawdown

One of the most popular risk indicators that can be easily applied to any automated trading strategy is maximum drawdown. Drawdown at moment T is defined as the difference between the current capital value $E(T)$ and the maximum capital value recorded during the preceding time period: $DD(T) = \max_{t<T} E(t) - E(T)$. Accordingly, the maximum drawdown corresponding to the time interval τ is $MaxDD = \max_{T \in \tau} DD(T)$.

The notion of drawdown is closely related to another indicator: the length of drawdown. This characteristic measures the time interval from local capital maximum to its breakdown. Let t_{max} be the moment of reaching the maximum capital value, and let $E(t_{max})$ be the value of capital at the moment t_{max}. If at the current moment in time the T capital value exceeds the previous maximum value, that is $E(T) > E(t_{max})$, the length of drawdown is fixed at $T - t_{max}$. The maximum drawdown length can be seen as an additional negative indicator reflecting the risk of the strategy.

Maximum drawdown and maximum drawdown length represent the most “emotional” risk characteristics. In real trading unacceptable values of these indicators often result in ceasing the utilization of otherwise profitable strategies. Meanwhile, periodical appearance of some drawdowns is a natural phenomenon for many successful strategies. It is important to note that the psychological effect produced by the drawdown depends on how successful the strategy was before the drawdown occurred. However, in terms of performance evaluation, high drawdown is undesirable regardless of the moment

when it emerged. At the same time, if the strategy had significant drawdown in the backtesting, it can theoretically take place just after the beginning of real trading. This can ruin the trading account beyond the recoverable level.

The main drawback of these indicators is that, while assessing the amount and length of maximum loss, they do not evaluate the probability of such an event. Meanwhile, the loss of a certain magnitude experienced by the strategy that uses just one year of historical data points to a much higher risk than the same loss recorded in the backtesting that is based on 10-year data. Hence, the riskiness of the strategy estimated on the basis of maximum drawdown should be weighted with respect to the testing period length.

Sharpe Coefficient

Since there is a direct positive relationship between return and risk, indicators that allow the simultaneous solution of two problems—return maximization and risk minimization—are very useful for performance evaluation. In the course of strategy backtesting, the developer obtains a sample of N return values. The closer the sample elements are to each other (and, hence, to their mean), the straighter and more stable is the equity curve. The efforts toward maximization of average return and simultaneous minimization of standard deviation can be realized by applying the Sharpe coefficient—an indicator which is widely used in almost all backtesting systems.

The Sharpe coefficient represents the ratio of average return to standard deviation. Although the return is usually reduced by a risk-free rate, we prefer to neglect it (using a risk-free rate complicates calculations without adding any supplementary benefit in assessing the strategy performance). In the exponential case (the strategy is based on reinvestment of profits generated during previous periods), we apply geometrical mean return r_e; in the linear case (the strategy is based on the investment of constant amounts), we apply arithmetical mean return r_l. The formula for mean-squared deviation is

$$\sigma = \frac{1}{N-1}\sqrt{\sum_{i=1}^{N}(r_i^d - r^d)^2}$$

where $d = l$ or $d = e$.

The main drawback of the Sharpe coefficient is that it does not account for the order in which profitable and unprofitable months (or other time frames) alternate. It follows from the Sharpe coefficient formula that we can shuffle summands in any order without changing

the result. This means that the same Sharpe coefficient can pertain to a strategy with evenly growing capital and to a strategy with an unacceptable maximum drawdown value. The strategy should not have long sequences of consecutive unprofitable months. This drawback can be compensated for by the joint use of the Sharpe coefficient and the previously mentioned risk indicators (particularly, maximum drawdown).

Profit/Loss Factor

The profit/loss factor, which is the ratio of the total profit of all profitable trades to the total loss of all unprofitable trades, is often used in the backtesting of trading strategies. There is a consensus that the profit/loss factor should exceed 2 for a strategy to be effective. This indicator is useful and informative when applied to strategies that execute single-type trades consecutively, one after another. For option strategies the calculation of the profit/loss factor has its specific aspects (since the set of trades generated by the strategy is not homogeneous).

Let us illustrate this by two examples. Consider the classical strategy of volatility trading with delta-neutral hedging. The simplest variant of this strategy involves buying (or selling) some option and iterative buying and selling of its underlying asset in different quantities (we will refer to the set of such trades as “the game”). There is no sense in considering profits and losses of separate trades for analyzing the performance of such strategy. Only the final result of the whole game, determined after closing all separate positions, can have sense. Thus, to calculate the profit/loss factor, one should use sums of all profitable and unprofitable games instead of using outcomes of separate trades.

The second example corresponds to the volatility selling strategy. Suppose that the strategy algorithm specifies the following sequence of procedures. Straddles, consisting of short at-the-money call and put options, are created every trading day for each underlying asset. The whole set of these combinations is sorted by the values of a specific criterion and a certain number of combinations is sold. Similarly to the previous example, evaluation of the profit/loss structure on the basis of separate trades is useless for such a strategy. It seems to be more appropriate to base the estimation of the profit/loss factor on profits and losses of separate option combinations.

To ensure proper application of the profit/loss factor, the developer should define the single event correctly (see section 5.5.1). In the first example the whole set of trades related to a specific underlying asset should be treated as a single event. In the second example the set of all trades executed within one day (or the whole expiration cycle) to

create a specific option combination should be regarded as a single event.

Consistency

The strategy is considered to be consistent if profitable and unprofitable trades are not concentrated within certain periods but are distributed more or less evenly through the whole testing period. Under uneven distribution, periods of capital growth alternate with periods of decline (their length and depth is determined by the extent of this unevenness). In option trading the precise sequence of trades following one after another is not always determinable. Usually positions that have been opened successively in time are not closed in the same order. Thus, for option strategies the consistency term should be associated with the equity curve (rather than with separate trades) that was established in the course of backtesting.

To express the consistency quantitatively, it would be practical to introduce the concept of "the ideal strategy." To be ideal, a strategy must have constant positive return and zero drawdowns. In the linear case (constant amounts are invested at each step), the equity curve of the ideal strategy is a straight line with the slope coefficient equal to the return. In the exponential case (money management is based on reinvestments), the equity curve is exponential. If a capital logarithm is used instead of absolute capital values, the equity curve is transformed into a straight line. The consistency indicator should measure the extent of deviation of real equity curve (with or without logarithmic transformation) from the ideal straight line.

Let the sequence $X = \{X(T),\ T = T0,\ T0 + 1,\ldots,\ TN\}$ represent the series of equities (or equity logarithms) that were successively measured during the simulation period τ. The straight line

$$y(T) = X(T_0) + \frac{X(T_N) - X(T_0)}{T_N - T_0}(T - T_0)$$

connects the first point of this sequence (when the trading starts) with the last one (the end of trading). The extent of deviation of the sequence X from the straight line can be estimated as the sum of squares:

$$\Phi(X, \tau) = \frac{1}{T_N - T_0 - 1} \sum_{T=T_0+1}^{T_N-1} (X(T) - y(T))^2 .$$

The lower the value of this indicator, the closer the equity curve is to the ideal line.

5.5.3. The Example of Option Strategy Backtesting

Let us consider performance evaluation of the strategy that is based

on selling SPY options and hedging these positions by long VIX options. The strategy employs the following algorithm: On some fixed day before the nearest expiration date, the SPY strangle is sold and the VIX call, belonging to the series with the next expiration date, is purchased in amounts determined by the capital management algorithm. The short position is bought back on a certain day before the expiration date. The long VIX option is held until expiration. These are the strategy parameters:

- The day of positions opening relative to the nearest expiration date
- The day of short options buyback
- The relative ratio of SPY strangles and VIX calls
- The share of capital invested in the current position
- The distance between strangles' strikes

Visual Analysis

Usually, analysis of backtesting results begins with the visualization of the time dynamics of basic performance indicators. The equity curve is one of the most informative visualizations. Figure 5.5.1 shows the equity curve obtained in the backtesting of one of the strategy variants. To demonstrate the desirable shape of the curve, we have selected the best variant of the strategy with rather smooth and stable capital growth. Visual analysis of the chart indicates that this strategy looks good, on the face of it, and deserves further detailed consideration.

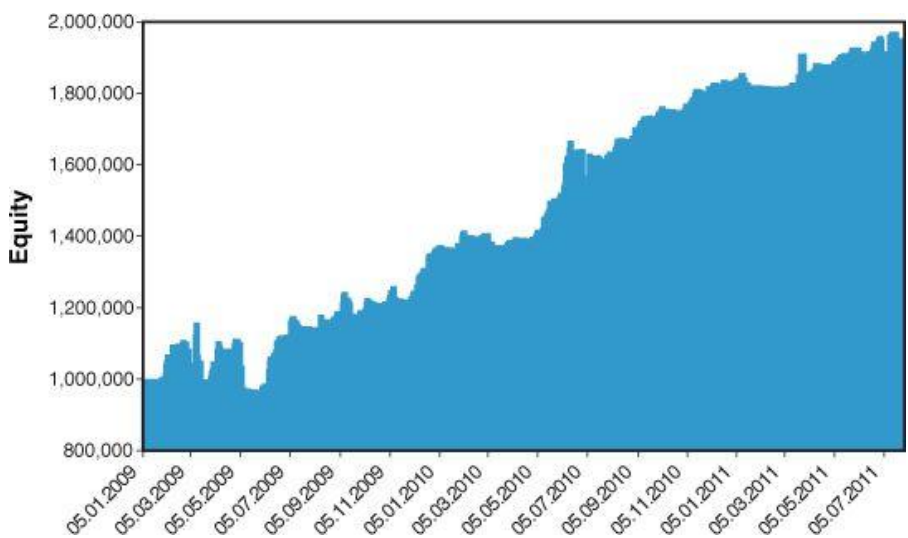

Figure 5.5.1. Equity curve obtained in the backtesting of one option strategy (see description in the text).

The equity curve is usually constructed using daily equity valuations (as shown in Figure 5.5.1). This allows intra-month evaluation of capital fluctuations and drawdowns. However, for the evaluation of option strategies, we also need to link performance to standard expiration cycles. A convenient form of such presentation is a chart showing monthly profits and losses (if the unit of time frame is one month). Figure 5.5.2 shows the monthly performance chart for the strategy used in our example. Such a presentation form enables instant performance overview—most of the 31 months, covered by the backtesting period, were profitable and only two months generated significant losses. Although these two strongly unprofitable months took place at the beginning of the testing period, they did not result in strategy failure, which is an important indicator of its effectiveness.

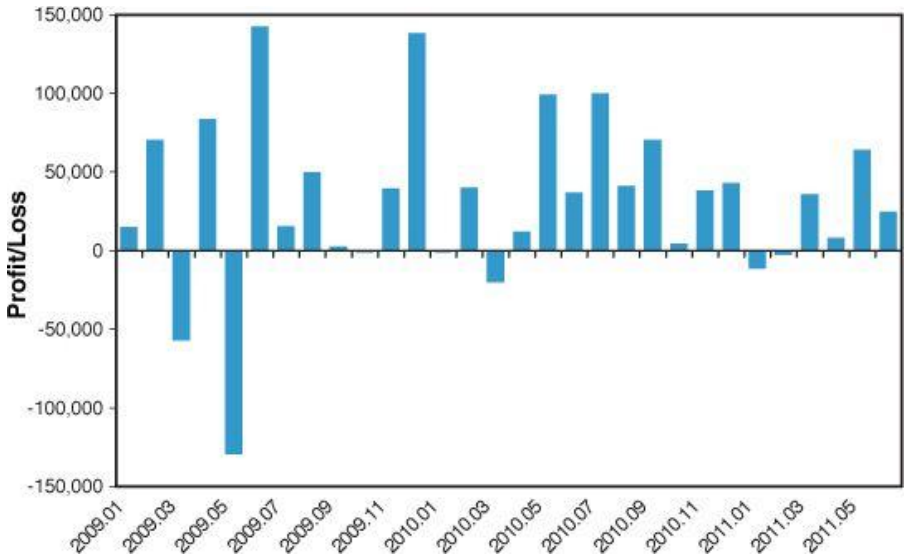

Figure 5.5.2. Monthly profits and losses obtained in the backtesting of the option strategy (see description in the text).

Analysis of Quantitative Characteristics

In the following text we present various performance indicators calculated for the linear version of the strategy (with a constant amount of capital invested at each trading step). Some indicators were calculated for two time frames: months (which corresponds to expiration cycles) and days (since day is the time frame at which steadiness of the strategy has to be controlled).

Testing period: 01.01.2009 to 31.07.2011.

Number of calendar days: $d = 938$.

Number of trading days: $t = 648$.

Number of months: $m = 31$.

Number of years: $n = \frac{938}{365.25} = 2.57$.

Starting capital: $E_0 = 1{,}000{,}000$.

Ending capital: $E_N = 1{,}953{,}594$.

Total profit: $P = 1{,}953{,}594 \text{ to } 1{,}000{,}000 = 953{,}594$.

Average annual profit: $\frac{P}{n} = \frac{953{,}594 \cdot 365.25}{938} = 371{,}322$.

Average monthly profit: $P_M = \frac{953{,}594}{31} = 30{,}761$.

Mean-squared deviation of monthly profits:

$$\sigma_M = \sqrt{\frac{\sum_{i=1}^{31} (P_{M_i} - 30{,}761)^2}{30}} = 53{,}563.$$

Sharpe coefficient for monthly data:

$$Sharpe_M = \sqrt{12} \cdot \frac{P_M}{\sigma_M} = \sqrt{12} \cdot \frac{30{,}761}{53{,}563} = 1.98.$$

Average daily profit: $P_D = \frac{953{,}594}{648} = 1{,}471.6$.

Mean-squared deviation of daily profits:

$$\sigma_D = \sqrt{\frac{\sum_{i=1}^{648} (P_{D_i} - 1{,}471.6)^2}{647}} = 21{,}119.$$

Sharpe coefficient for daily data:

$$Sharpe_D = \sqrt{252} \cdot \frac{P_D}{\sigma_D} = \sqrt{252} \cdot \frac{1{,}474.6}{21{,}119} = 1.1.$$

Maximum drawdown occurred on 18.03.2009, when the capital had decreased from its maximum of 1,157,537 (recorded on 11.03.2009) to 818,733. The drawdown amount was 338,803.

The length of maximum drawdown was 78 days. It turned out to be the most prolonged drawdown recorded during the entire testing period.

Percentage of profitable transactions (the transaction is defined as a set of all trades relating to a certain combination and executed

within one day): 53.6%.

Percentage of profitable months: 76.6%.

Maximum number of consecutive profitable months: 9 months.

Maximum number of consecutive unprofitable months: 2 months.

Average number of consecutive profitable months: 3.3 months.

Average number of consecutive unprofitable months: 1.2 months.

Total profit of profitable trades: 2,684,032.

Total loss of unprofitable trades: 1,730,438.

Profit/Loss factor, calculated on the basis of separate trades: 1.55.

For this strategy, as well as for many other option strategies, the profit/loss factor calculated on the basis of separate trades is meaningless (see the explanation in section 5.5.2). A similar ratio calculated using monthly profits and losses is more informative. Such an approach is quite natural for the strategy that was considered in our example. The profit/loss factor calculated on the basis of monthly data is significantly higher than the same ratio calculated using separate trades: 1,176,797 / 223,203 = 5.3.

5.6. Establishing the Backtesting System: Challenges and Compromises

The first challenge faced by the system developer consists in maintaining the historical database that contains complete information required for adequate testing of option trading strategies. Information must be reliable and free of inaccuracies and errors. These two requirements to the database—completeness and reliability—contradict each other to some extent. By striving to include as much information as possible into the database, the developer inevitably faces the problem of information reliability. The more information of different types that is accumulated, the higher the probability of including some erroneous or inaccurate data in the database. Therefore, efforts applied to maximize the completeness of information should be balanced by the possibility of its verification.

Filtration of unacceptable signals is a compromise between striving for maximum filtration austerity, on the one hand, and the desirability of avoiding elimination of potentially profitable trading variants, on the other. The stricter the filtration, the higher the probability of discarding promising trading opportunities. Simulating prices and volumes of signal executions is also based on compromise. The more

conservative the approach used in the simulation (executed volumes and prices are assumed to be worse than those used for signals generation), the lower the estimated strategy effectiveness and the higher the probability that the results of real trading will not be worse than backtesting results.

One of the most critical issues in developing a backtesting system is the division of the historical period covered by the database into in-sample and out-of-sample periods. The longer the in-sample period, the shorter the out-of-sample period (and vice versa). In general, one can claim that the developer should strive for prolonging the out-of-sample period as much as possible. This allows testing the strategies at various market phases (using data not involved in strategy optimization). However, extending the out-of-sample period inevitably results in shortening the in-sample period. As a result, the optimization will not be sufficiently reliable (since the in-sample period will not include all possible market phases).

The most challenging problem that arises in the backtesting of all trading strategies (and especially of the option-based strategies) is the danger of overfitting. This problem becomes worse when the number of parameters used in strategy creation increases. Accordingly, the probability of overfitting can be reduced by decreasing the number of optimized parameters. However, a reasonable compromise is needed here, since excessive reduction in the number of parameters may decrease strategy flexibility. As a result, a potentially profitable version of the strategy may be left out.

Bibliography

Aldridge, Irene. *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems.* Wiley, 2009.

Bandy, Howard B. *Quantitative Trading Systems.* Blue Owl Press, 2007.

Barmish, B. Ross. On trading of equities: a robust control paradigm. *Proceedings of the 17th IFAC World Congress,* pp. 1621-1626, 2008.

Briza, Antonio C., and Prospero C. Naval Jr. Design of stock trading system for historical market data using multiobjective particle swarm optimization of technical indicators. *Proceedings of GECCO 2008.* ACM, pp. 1871-1878, 2008.

Bryant, Michael R. Building trading system using automatic code generation. Working paper, 2010.

Burns, Patrick. Random portfolios for evaluating trading strategies. Working paper, 2006.

Chan, Ernest P. *Quantitative Trading: How to Build Your Own Algorithmic Trading Business.* Wiley, 2008.

Harris, Larry. *Trading and Exchanges: Market Microstructure for Practitioners*. Oxford University Press, 2002.

Harris, Michael. *Profitability and Systematic Trading: A Quantitative Approach to Profitability, Risk, and Money Management.* Wiley, 2008.

Izraylevich, Sergey, and Vadim Tsudikman. *Systematic Options Trading: Evaluating, Analyzing, and Profiting from Mispriced Option Opportunities.* FT Press, 2010.

Izraylevich, Sergey, and Vadim Tsudikman. When volatility distorts probability. *Futures*, vol. 40 (1), pp. 44-47, 2011.

Izraylevich, Sergey, and Vadim Tsudikman. Measuring new risk management tools. *Futures*, vol. 39 (1), pp. 42-47, 2010.

Izraylevich, Sergey, and Vadim Tsudikman. Risk management: facing new challenges. *Futures*, vol. 38 (12), pp. 30-35, 2009.

Izraylevich, Sergey, and Vadim Tsudikman. Short volatility trading in extreme markets. *Futures*, vol. 38 (10), pp. 28-31, 2009.

Izraylevich, Sergey, and Vadim Tsudikman. A better risk gauge for options portfolios. *Futures*, vol. 38 (7), pp. 24-27, 2009.

Izraylevich, Sergey, and Vadim Tsudikman. An empirical

solution to option pricing. *Futures*, vol. 38 (5), pp. 28-31, 2009.

Kan, Raymond, and George Kirikos. Biases in evaluating trading strategies. Working paper, 1995.

Katz, Jeffrey Owen, and Donna L. McCormick. *The Encyclopedia of Trading Strategies.* McGraw-Hill, 2000.

Kim, Kendall. *Electronic and Algorithmic Trading Technology: The Complete Guide.* Academic Press, 2007.

Masters, Timothy. Monte-Carlo evaluation of trading systems. Working paper, 2006.

Meyers, Dennis. The siren call of optimized trading systems. Working paper, 1996.

Moody, John, Lizhong Wu, Yuansong Liao, and Matthew Saffell. Performance functions and reinforcement learning for trading systems and portfolios. *Journal of Forecasting*, vol. 17, pp. 441-470, 1998.

Munir, Danish, and Divya Krishnan. Financial modeling: a system to test high-frequency trading strategies. Working paper, 2009.

Narang, Rishi K. *Inside the Black Box: The Simple Truth About Quantitative Trading.* Wiley, 2009.

Owens, Scott. Historical testing. Working paper, 2005.

Oya, Tomoyasu. Evaluating automated trading systems using random walk securities. *Proceedings of the 2008 International Conference on Genetic and Evolutionary Methods.* CSREA Press, pp. 174-178, 2008.

Pardo, Robert. *The Evaluation and Optimization of Trading Strategies.* Wiley, 2008.

Pavlidis, N.G., E.G. Pavlidis, M.G. Epitropakis, V.P. Plagianakos, and M.N. Vrahatis. Computational intelligence algorithms for risk-adjusted trading strategies. Working paper, 2006.

Pinsky, E., and R. Sunitsky. Testing a high frequency trading strategy. Working paper, 2010.

Ruggiero, Murray A. *Cybernetic Trading Strategies. Developing a Profitable Trading System with State-of-the-Art Technologies.* John Wiley & Sons, 1997.

Schoenberg, Ronald. Using DOE method in trading. *Futures*, vol. 39 (2), 2010.

Schroeder, Michael, Julie McCann, and Daniel Haynes. Trading

without explicit ontologies. *Proceedings of Agent-mediated E-commerce.* Springer-Verlag, 1999.

Stridsmam, Thomas. *Trading Systems and Money Management.* McGraw-Hill, 2003.

Tsudikman, Vadim, and Sergey Izraylevich. Best method of multi-criteria analysis. *Futures*, vol. 39 (6), pp. 40-42, 2010.

Tsudikman, Vadim, and Sergey Izraylevich. Multi-criteria analysis: A practical approach. *Futures*, vol. 39 (7), pp. 25-27, 2010.

Tsudikman, Vadim, Sergey Izraylevich, and Arsen Balishyan. *Risk Evaluation and Financial Crises: Review of New Approaches to Risk Evaluation: VaR Criticism, Alternatives and Modifications*. FT Press, 2011.

Vanstone, Bruce J., and Gavin Finnie. *An Empirical Methodology for Developing Stock-market Trading Systems Using Artificial Neural Networks.* Bond University, 2007.

Vince, Ralph. *The Mathematics of Money Management: Risk Analysis Techniques for Traders.* Wiley, 1992.

Vliet, Benjamin Van. *Building Automated Trading Systems.* Elsevier Inc, 2007.

Appendix

Basic Notions

Call option. A standard contract which gives its buyer the right (without imposing any obligations) to buy a certain *underlying asset* at a specific point in time in the future (see *expiration date*) for a fixed price (see *strike price*).

Put option. A standard contract which gives its buyer the right (without imposing any obligations) to sell a certain *underlying asset* at a specific point in time in the future (see *expiration date*) for a fixed price (see *strike price*).

Strike price. The price, settled in a contract, at which the option buyer can realize his right to buy or sell the underlying asset.

Expiration date. The date, settled in a contract, before which (see *American option*) or at which (see *European option*) the option buyer can realize his right to buy or sell the underlying asset.

American option. A type of the option contract which gives its buyer the right to buy or sell the underlying asset at any point until the *expiration date*.

European option. A type of the option contract which gives its buyer the right to buy or sell the underlying asset at the *expiration date*.

Premium. The value of the option paid by the buyer to the seller. The premium of stock options is usually quoted for 1 share, but the standard contract is generally comprised of options for 100 shares.

Intrinsic value. A part of the option *premium* which is equal (for the call option) to the difference between the current underlying asset price and the *strike price* if this difference is positive, and zero otherwise. For the put option the intrinsic value is equal to the difference between the *strike price* and the current underlying asset price if this difference is positive, and zero if it is negative.

Time value. Another part of the option *premium* which is equal to the difference between the option premium and the *intrinsic value*. Time value is a risk premium paid by option buyers to sellers.

Time decay. The process of options *time value* decreasing while the *expiration date* approaches. The closer the expiration date, the faster the decay.

Options in-the-money. Options with the strike price distant from the

current underlying asset price. For put options the strike price is higher, and for call options it is lower than the underlying asset price. The *intrinsic value* of such options is high, whereas their *time value* is relatively low.

Options out-of-the-money. The strike price of these options is also distant from the current underlying asset price. However, in this case the strike of put options is lower than the underlying asset price (and vice versa for call options). The *intrinsic value* of such options is zero, and their *time value* is relatively low.

Options at-the-money. Options with the strike price close to the current underlying asset price. The *intrinsic value* of such options is low, while their *time value* is relatively high.

Combination. Any number of different options corresponding to the same underlying asset can be considered and analyzed as a whole entity. Options comprising a combination may be long (bought) and/or short (sold), and may have the same or different expiration dates and strike prices.

Margin requirements. The minimum amount of capital required to maintain an option position. It is set by the regulating authorities and is generally made tougher by the exchange and/or broker. The amount of margin requirements is recalculated daily in accordance with the underlying asset price change.

Historical volatility. The indicator reflecting the extent of the underlying asset price variability at the specific time interval. It is usually estimated as the standard deviation of relative increments of the underlying asset prices normalized by the square root of time. This method is straightforward and the simplest one, although there are a lot of other more elaborate and sophisticated techniques. Historical volatility is widely used as the basic risk indicator. In option theory it is a key element of all pricing models.

Implied volatility. Another indicator reflecting the underlying asset price variability. Unlike the *historical volatility,* it is not based on historical prices of the underlying asset, but is derived from current market prices of the option. Pricing models generally use the *historical volatility* value to calculate the theoretical option fair value. If instead we substitute the option market price in the model formula and solve the inverse problem, we can derive volatility "implied" by the option price. Implied volatility expresses market expectations for future variability of the underlying asset price. Besides, this indicator is a convenient measure of option expensiveness. The divergence of implied and historical volatility often creates favorable opportunities for implementing different option trading strategies.

The Greeks (vega, gamma, delta, rho, theta). Indicators used to estimate the risks of options reflecting the extent to which the value of the option responds to small changes in specific variables. The underlying asset price, volatility, time, and interest rate can be taken as variables. *The Greeks* are calculated analytically as partial derivatives of the option price with respect to a given variable. To calculate a derivative, a certain option pricing model is used (for example, the Black-Scholes formula). These indicators can be interpreted as the speed of the option price change in response to the variable change.

Delta. The first partial derivative of the option price with respect to the underlying asset price. It estimates the change in the option price given a one-point change in the underlying asset price. The delta is always positive for a call option and negative for a put option.

Gamma. The second partial derivative of the option price with respect to the underlying asset price. It estimates the change in the delta given a one-point change in the underlying asset price. Gamma shows if the speed of the delta change is increasing, decreasing, or constant.

Vega. The first partial derivative of the option price with respect to the underlying asset volatility. It estimates the change in the option price given a one-point change in the underlying volatility. This indicator plays an important role in the strategies of volatility trading.

Theta. The first partial derivative of the option price with respect to time. It estimates the change in the option price in response to the decrease in time left until expiration by one unit. Theta characterizes the speed of *time decay*.

Rho. The first order derivative of the option price with respect to the risk-free interest rate. It estimates the change in the option price given a one-point change in the interest rate. *Rho* expresses the sensitivity of the option price to changes in the interest rate.

Payoff function. The relationship between the price of a separate option or a combination and the underlying asset price, estimated for a specific future date. The payoff function of a combination consisting of options with the same expiration date and calculated for the expiration date is a broken line. If the estimation is made for the date preceding expiration or if the combination includes options with different expiration dates, the payoff function is a curve (calculated on the basis of a specific option pricing model).

Payoff Functions

Separate Options

Figure A.1 shows payoff functions of long and short call and put options. Solid lines depict payoff functions for the expiration date; dashed lines, for some date before the expiration. At a given underlying asset price, the more time left until expiration, the higher the value of the long position (in the figures the dashed lines of both call and put options are higher than solid ones). Accordingly, the closer the expiration date, the higher the value of the short position (dashed lines are lower than solid ones in the figures). The reason is that the option value consists of *intrinsic* and *time values*. The latter decreases continuously when approaching the expiration date (this phenomenon is called *time decay*).

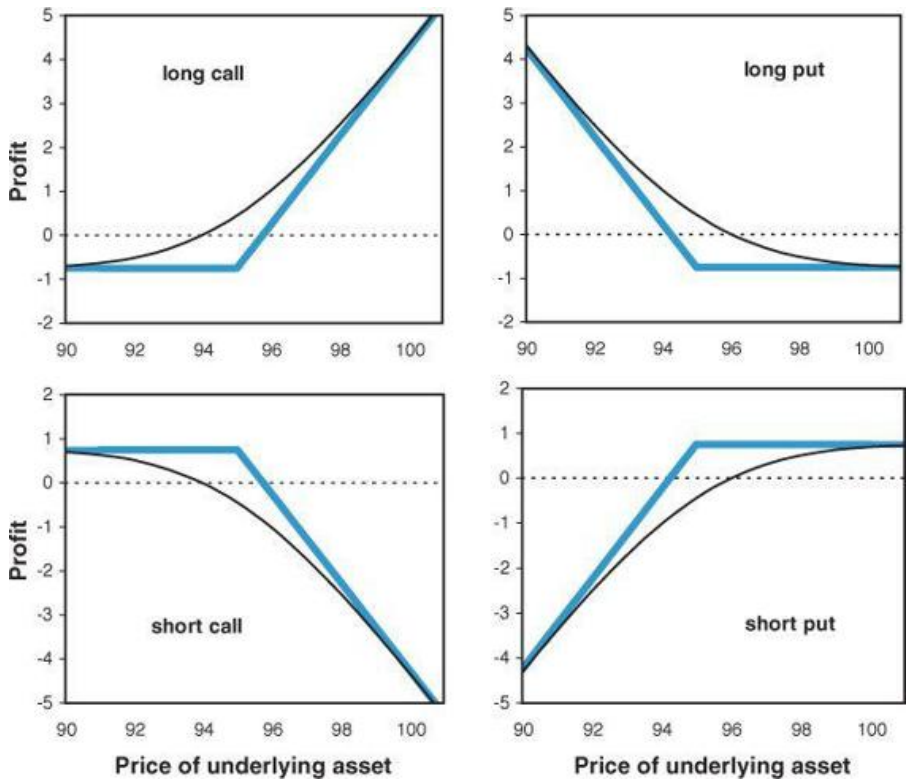

Figure A.1. Payoff functions reflecting profits and losses of buyers (long) and sellers (short) of call and put options. Bold lines show payoff functions corresponding to the expiration date; thin lines, to a date before the expiration.

The buyer of a call option realizes theoretically unlimited profit when the underlying asset price grows and limited loss when it decreases. The put option buyer realizes theoretically unlimited profit when the underlying asset price decreases and limited profit when it increases. In both cases the maximum loss is limited to the premium

paid.

Payoff functions of short options are the opposite of those of long options (see Figure A.1). Profits of the seller are limited to the premium received, and the losses are theoretically unlimited and depend on the extent of underlying asset price move. The more the underlying asset price increases, the higher the loss of the short call is. Correspondingly, the lower the price decrease, the higher is the loss of the short put.

Option Combinations

More complicated forms of payoff functions can be obtained when several different options are combined. This feature represents one of the most important advantages of investing in options. Construction of complex option combinations allows creating a multitude of nonlinear payoff functions, thereby considerably enlarging the potentialities and flexibility of the sophisticated investor.

Applying different principles of options combining (such as ratio of long to short options, calls to puts ratio, positioning strike prices relative to the current underlying price, selection of expiration dates, and so on), one can create combinations with almost any desired type of payoff function. The combinations are usually classified by the principles of their creation and the shape of payoff functions characteristic of them. In the literature (including this book) such classes are called "option strategies." Next we describe several examples of the most popular strategies.

Straddle

The *long straddle* is created by buying call and put options with the same strike price and expiration date. The number of calls is usually equal to the number of puts (as in Figure A.2), but this is not a compulsory condition. If the underlying asset price is close to the strike price at the expiration date, the combination is unprofitable with a loss limited to the premium paid for it. If a considerable price movement occurs (regardless of the direction), the straddle generates profit. The *short straddle* is created under the same principles but the options are sold rather than bought. Accordingly, the combination is profitable if the underlying asset price does not change a lot and generates losses when there is a considerable price movement in any direction (see Figure A.2).

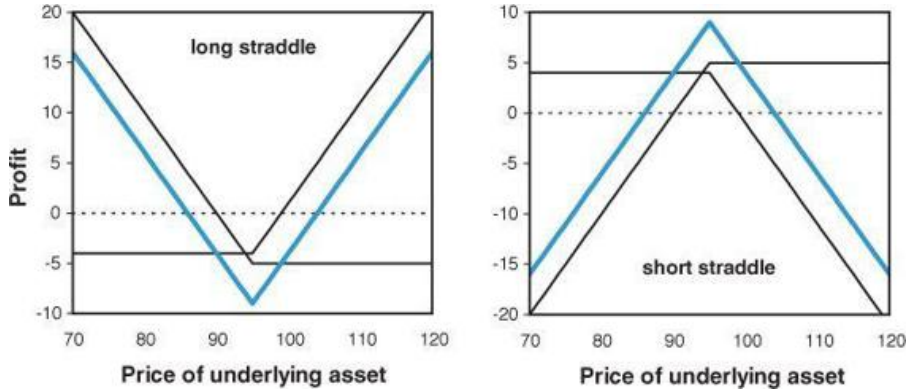

Figure A.2. Payoff functions of long and short straddles corresponding to the expiration date (bold lines). Thin lines show payoff functions of separate options forming the combinations.

Strangle

The *long strangle* is created by buying call and put options with the same expiration date but different strike prices. The strike of the call is usually higher than that of the put, and their quantities in the combination are equal, although neither of these conditions is compulsory. Correspondingly, the *short strangle* combination is created by selling these options. Payoff functions of long and short strangles are shown in Figure A.3. Profits and losses of strangles depend on the changes of the underlying asset price in the same manner as was described for straddles. The difference between these two strategies is that the maximum loss of long strangles and the maximum profit of short strangles are lower than those of straddles. However, the probabilities of these profits and losses are higher for strangles than for straddles.

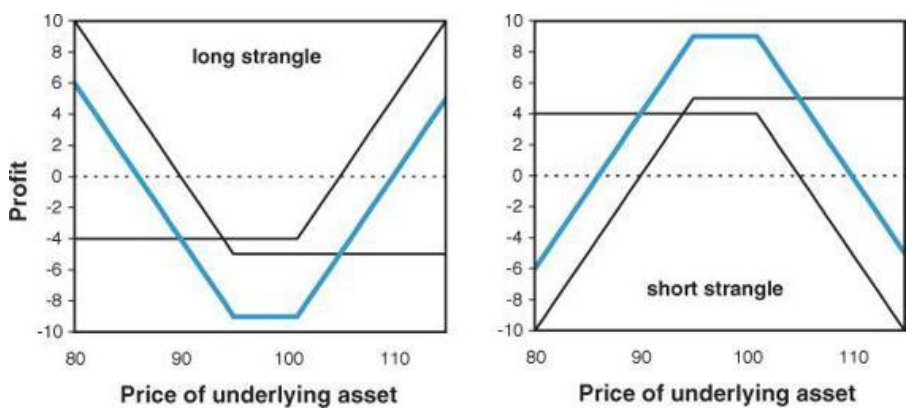

Figure A.3. Payoff functions of long and short strangles corresponding to the expiration date (bold lines). Thin lines show payoff functions of separate options forming the combinations.

Straddles and *strangles,* discussed previously, are created by combining options with the same expiration date. Calendar spreads consist of options with different expiration dates. By selling the call (or put) option with the nearest expiration date and buying the call (or put) option with the later expiration date, we construct the combination corresponding to the *short calendar spread* strategy. This is a debit strategy (i.e., it requires investing funds) since the option bought is always more expensive than the option sold (owing to the fact that the premium of more distant option has more *time value*). The payoff function shown in Figure A.4 is calculated for the expiration date of the nearest option (we assume closing the second option position on that date). This strategy generates a limited profit if the underlying asset price changes are small. If price moves are substantial, the combination generates a limited loss. Strike prices of both options may be the same (as in Figure A.4) or different. In the latter case the amount of the maximum possible profit is lower but its probability is higher.

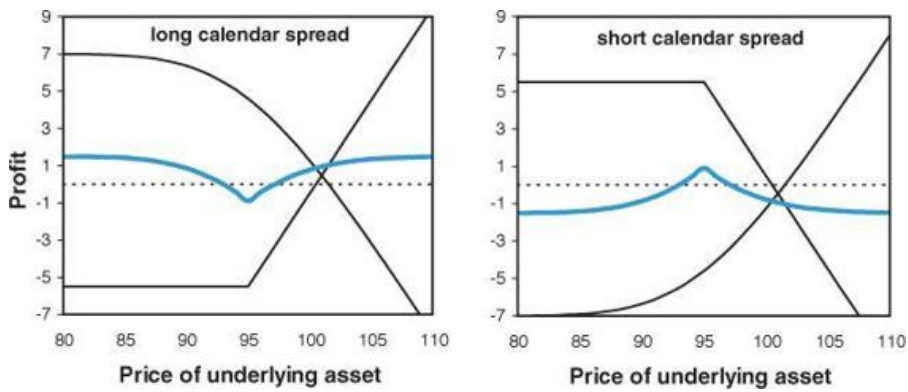

Figure A.4. Payoff functions of short and long calendar spreads. Bold lines show payoff functions of combinations for the expiration date of the nearest option, and thin lines show payoff functions of separate options forming the combinations (broken lines denote options with the nearest expiration date, and curves denote options with a later expiration date).

The *long calendar spread* strategy is the opposite of the *short calendar spread* in every respect. The combination is created by selling the option with the nearest expiration date and buying the option with the later expiration date. Accordingly, this strategy is a credit one. Both profits and losses of this combination are limited. Profit is generated under substantial price movements, and loss is generated when the underlying asset price does not change considerably (see Figure A.4).

All combinations described previously are related to market-neutral

strategies. It means that short combinations become profitable when the underlying asset price fluctuates within a quite narrow range. Losses are generated as a consequence of strong price movements. The opposite is true for long combinations. Such strategies are called neutral since they do not require forecasting the market direction (i.e., growth or fall of the underlying asset price). Although there are a lot of market-neutral strategies, we restrict our description to those which are mentioned most often in this book. Besides market-neutral strategies, there is an extensive class of option strategies based on forecasting future price direction. But since this book is dedicated mostly to neutral strategies, we will discuss only a few combinations relating to such strategies (more detailed descriptions can be easily found in the literature).

Bull and Bear Spreads

Buying a call option with a certain strike price and selling a call option with a higher strike price produces a *bull spread.* This strategy makes a limited profit when the underlying asset price increases, and yields a limited loss if it decreases (see Figure A.5). Both options have the same expiration date. Since the call premium is higher for the lower strike prices, a *bull spread* is a debit combination requiring initial investments. A *bull spread* can also be created by buying a put with a lower strike price and selling a put with a higher strike price (in this case it will be a credit combination).

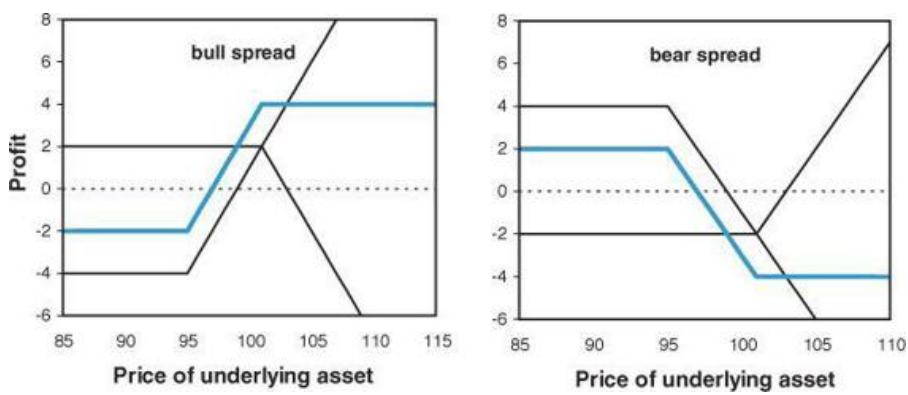

Figure A.5. Payoff functions of bull and bear spreads corresponding to the expiration date (bold lines). Thin lines show payoff functions of separate options forming the combinations.

A *bear spread* is created by buying a call (put) with a certain strike price and selling a call (put) with a lower strike price. Such a combination generates a limited profit if the underlying asset price falls, and a limited loss otherwise (see Figure A.5). Using calls allows creating a credit position, whereas a debit position is created using put

options.

Index

A

acceptable values, determining range of, 76, 85-87
access in historical database (in backtesting systems), 220-221
adaptive optimization, 233-234
additive convolution, 97, 172
adjacent cells, average of, 103-104
alternating-variable ascent optimization method, 116-118
    comparison with random search method, 132
    effectiveness of, 128
American option, defined, 251
amoeba search optimization method, 123-127
analysis of variance (ANOVA) in one-dimensional capital allocation system, 190-191
analytical method (index delta calculation), 142-143
analytical method (VaR calculation), 137
arbitrage situations, tests for, 223
assets
    in delta-neutral strategy, 5
    in portfolio
        analysis of delta-neutrality strategy, 24-25
        analysis of partially directional strategies, 58
    price forecasts, 35
        call-to-put ratio in portfolio, 40-49
        delta-neutrality applied to, 49-57
        embedding in strategy structure, 36-40
asymmetry coefficient, 157-159, 180-181
asymmetry of portfolio
    analysis of delta-neutrality strategy, 29-30
    analysis of partially directional strategies, 60
at-the-money, defined, 252
attainability of delta-neutrality, 14, 19-20, 51, 54-55

automated trading systems
defined, xv
empirical approach to development, xviii-xix
rational approach to development, xix-xx
scientific approach to development, xviii
average of adjacent cells, 103-104

B

backtesting systems
challenges and compromises, 246
framework for, 232
adaptive optimization, 233-234
in-sample optimization/out-of-sample testing, 232-233
overfitting problem, 234-236
historical database, 217
data access, 220-221
data reliability/validity, 222-224
data vendors for, 218
recurrent calculations, 221-222
structure of, 219-220
order execution simulation, 228
commissions, 231
price modeling, 230-231
volume modeling, 229
performance evaluation indicators, 236
backtesting example, 242-245
characteristics of return, 237-238
consistency, 241
maximum drawdown, 238-239
profit/loss factor, 240-241
Sharpe coefficient, 239-240
single events, 236
unit of time frame, 236
signals generation, 225

filtration of signals, 227-228
functionals development/evaluation, 226-227
principles of, 225-226
bear spreads, payoff functions, 258-259
boundaries of delta-neutrality, 6-10
in calm versus volatile markets, 10-11, 13
optimal portfolio selection, 67-72
in partially directional strategies, 49-55, 57
portfolio structure and properties at, 62-65
quantitative characteristics of, 14-21
broker commissions (in backtesting systems), 231
bull spreads, payoff functions, 258-259

C

calendar optimization, defined, 73
calendar spreads, payoff functions, 257-258
call options
defined, 251
payoff functions, 254-255
call-to-put ratio in portfolio, 40-42
factors affecting, 44-49
calm markets
delta-neutrality boundaries in, 10-13, 51-52
portfolio structure analysis
long and short combinations, 26-27
loss probability, 31-33
number of combinations in portfolio, 22-24
number of underlying assets in portfolio, 24-25
portfolio asymmetry, 29-30
straddles and strangles, 28-29
VaR, 33-34
capital allocation
challenges and compromises, 214-216
classical portfolio theory, 168-169

option portfolios, features of, 169-170
in delta-neutral strategy, 5
indicators
asymmetry coefficient, 180-181
delta, 180
expected profit, 179
inversely-to-the-premium, 175-176
inversely-to-the-premium versus stock-equivalency, 176-178
profit probability, 179
stock-equivalency, 174-175
VaR, 181-183
weight function for return/risk evaluation, 178-179
multidimensional system, 172, 204-205
one-dimensional system versus, 206-209
one-dimensional system, 170, 172
analysis of variance in, 190-191
factors affecting, 183-186
historical volatility in, 186-187
measuring capital concentration, 192-195
multidimensional system versus, 206-209
number of days to expiration in, 187-188
number of underlying assets in, 188-190
weight function transformation, 196-204
in partially directional strategies, 43
portfolio system, 209-211
elemental approach versus, 173-174, 211-214
capital concentration
concave versus convex weight function comparison, 202-204
measuring, 192-195
one-dimensional versus multidimensional capital allocation systems, 208-209
portfolio versus elemental capital allocation systems, 213-214

capital management systems
first level of, 167
in partially directional strategies, 43
second level of, 167
characteristics of return performance indicator, 237-238
classical portfolio theory, 168-169
option portfolios, features of, 169-170
closing signals
in delta-neutral strategy, 4
generating (in backtesting systems), 225
filtration of signals, 227-228
functionals development/evaluation, 226-227
principles of, 225-226
in partially directional strategies, 42
combination of options, 3
combinations
defined, 252
factors affecting call-to-put ratio, 44-49
long and short combinations, analysis of delta-neutrality strategy, 26-27
in partially directional strategies, 43
payoff functions for, 255
bull/bear spreads, 258-259
calendar spreads, 257-258
straddles, 256
strangles, 256-257
in portfolio
analysis of delta-neutrality strategy, 22-24
analysis of partially directional strategies, 57-59
commissions (in backtesting systems), 231
computation, defined, 74
concave weight function, 196, 198
convex weight function compared
by capital concentration, 202-204

by profit, 200-202
conditional optimization, 74
consistency performance indicator, 241
constraints in conditional optimization, 74
convex weight function, 196-198
concave weight function compared
by capital concentration, 202-204
by profit, 200-202
convolution
of indicators, 172
in multicriteria optimization, 97-98
correlation analysis
of objective functions, 91-96
risk indicator interrelationships, 162-165
correlation coefficient in objective function relationships, 91
criterion parameters in partially directional strategies, 42
criterion threshold index, 14-17, 51-52
point of delta-neutrality, determining, 7-8

D

data access in historical database (in backtesting systems), 220-221
data reliability in historical database (in backtesting systems), 222-224
data validity in historical database (in backtesting systems), 222-224
data vendors for historical database (in backtesting systems), 218
database (in backtesting systems), 217
data access, 220-221
data reliability/validity, 222-224
data vendors for, 218
recurrent calculations, 221-222
structure of, 219-220
deformable polyhedron optimization method, 123-127

delta, 138-139, 169, 180, 253. *See also* index delta
delta-neutral strategy, xvii, 4
basic form of, 4-5
optimal portfolio selection, 67-72
optimization space of, 79-80
acceptable range of parameter values, 85-87
optimization dimensionality, 80-85
optimization step, 87-88
partially directional strategies versus, 34
points and boundaries of, 6-10
in calm versus volatile markets, 10-11, 13
quantitative characteristics of, 14-21
portfolio structure analysis, 21-34
long and short combinations, 26-27
loss probability, 31-33
number of combinations in portfolio, 22-24
number of underlying assets in portfolio, 24-25
portfolio asymmetry, 29-30
straddles and strangles, 28-29
VaR, 33-34
portfolio structure and properties at boundaries, 62-65
price forecasts versus, 35
delta-neutrality
applied to partially directional strategies, 49-55, 57
attainability, 14, 19-20, 51, 54-55
derivatives, Greeks as, 138
determination coefficient in objective function relationships, 91
dimensionality of optimization, 80-85
one-dimensional optimization, 80-82
two-dimensional optimization, 83-85
direct filtration method, 227
direct methods, defined, 78
direct search optimization methods, 115
alternating-variable ascent method, 116-118

comparison of effectiveness, 127-130
drawbacks to, 115-116
Hook-Jeeves method, 118-120
Nelder-Mead method, 123-127
Rosenbrock method, 120-123
directional strategies, xvi
diversification, underlying assets in portfolio, 24
diversity of options available, 3
domination in Pareto method, 99

E

elemental capital allocation approach, 173-174
portfolio system versus, 211-214
embedding price forecasts in strategy structure, 36-40
empirical approach to automated trading system development, xviii-xix
equity curve in backtesting results, 242-244
European option, defined, 251
evaluation
of option pricing, 1-2
of performance (in backtesting systems), 236
backtesting example, 242-245
characteristics of return, 237-238
consistency, 241
maximum drawdown, 238-239
profit/loss factor, 240-241
Sharpe coefficient, 239-240
single events, 236
unit of time frame, 236
exhaustive search optimization method, 114-115
expected profit (capital allocation indicator), 179
expiration date, 169
defined, 251
in delta-neutral strategy, 8-13
effect on call-to-put ratio, 48-49

in one-dimensional capital allocation system, 187-188
exponential annual return, 237

F

fair value pricing, determining, 1-2
filtration of signals, 227-228
first level of capital management systems, 167
fixed parameters, steadiness of optimization space, 109-110
forecasts of underlying asset prices, 35
call-to-put ratio in portfolio, 40-42
factors affecting, 44-49
delta-neutrality applied to, 49-57
embedding in strategy structure, 36-40
full optimization cycle, defined, 74
functionals, development and evaluation of, 226-227
fundamental analysis for price forecasts, 35

G

gamma, 138, 253
generating signals in delta-neutral strategy, 4
genetic algorithms, 115
global maximum, defined, 75
Greeks, 169
defined, 253
in risk evaluation, 138-139

H

hedging strategies, xvi
historical database (in backtesting systems), 217
data access, 220-221
data reliability/validity, 222-224
data vendors for, 218
recurrent calculations, 221-222
structure of, 219-220
historical method (VaR calculation), 137

historical optimization period, steadiness of optimization space, 112-114
historical volatility
defined, 252
in delta-neutral strategy, 80
in one-dimensional capital allocation system, 186-187
recurrent calculations applied, 221-222
in risk evaluation, 136
Hook-Jeeves optimization method, 118-120
comparison with random search method, 132
effectiveness of, 128

I–J–K

ideal strategy, 241
implied volatility
defined, 252
estimations, 224
in-sample optimization, 232-233
in-the-money, defined, 252
index delta, 139, 141, 169
analysis of effectiveness
at different time horizons, 150-156
in risk evaluation, 146-149
analytical method of calculation, 142-143
applicability of, 156-157
calculation algorithm, 141-142
example of calculation, 144-146
indirect filtration method, 227
interrelationships
between risk indicators, 161
correlation analysis, 162-165
testing, 161
of objective functions, 91-96
intrinsic value, defined, 251
inversely-to-the-premium (capital allocation indicator),

175-176
stock-equivalency versus, 176-178
investment assets
in delta-neutral strategy, 5
in portfolio
analysis of delta-neutrality strategy, 24-25
analysis of partially directional strategies, 58
price forecasts, 35
call-to-put ratio in portfolio, 40-49
delta-neutrality applied to, 49-57
embedding in strategy structure, 36-40
isolines in delta-neutrality boundaries, 9

L

length of delta-neutrality boundary, 14, 17-19, 51-54
life span of options, limited nature of, 2-3
linear annual return, 237
linear assets, options versus, xvii
linear financial instruments
nonlinear instruments versus, 135
risk evaluation, 136-137
local maximum, defined, 75
long calendar spreads, 258
long combinations
analysis of delta-neutrality strategy, 26-27
factors affecting call-to-put ratio, 44-49
in portfolio, analysis of partially directional strategies, 58-59
long options, payoff functions, 254-255
long positions, limitations on, 111
long straddles, 256
long strangles, 256
loss probability, 159-160
analysis of delta-neutrality strategy, 31-33
in portfolio, analysis of partially directional strategies, 60

M

margin requirements, 170, 252
market impact, 230
market volatility, effect on call-to-put ratio, 46-47
market-neutral strategies, 258. *See also* delta-neutral strategy
market-neutrality, xvii
Markowitz, Harry, 168
maximum drawdown, 238-239
    as objective function
        effect on optimization space, 90
        relationship with percentage of profitable trades, 93
        relationship with profit, 92
mean, ratio to standard error, 104-105
minimax convolution, 97
modeling (in backtesting systems), 228
    commissions, 231
    price modeling, 230-231
    volume modeling, 229
money management in delta-neutral strategy, 5
Monte-Carlo method (VaR calculation), 137
moving averages, compared with average of adjacent cells, 103
multicriteria optimization, 79
    convolution, 97-98
    nontransitivity problem, 96-97
    Pareto method, 99-102
    robustness of optimal solution, 102
        averaging adjacent cells, 103-104
        ratio of mean to standard error, 104-105
        surface geometry, 106-108
    steadiness of optimization space, 108-109
        relative to fixed parameters, 109-110
        relative to historical optimization period, 112-114
        relative to structural changes, 110-111
multidimensional capital allocation system, 172, 204-205

one-dimensional system versus, 206-209
multiple regression analysis in one-dimensional capital allocation system, 190-191
multiplicative convolution, 97, 172

N

Nelder-Mead optimization method, 123-129
neuronets, 115
nodes, defined, 74
nonlinear financial instruments
linear instruments versus, 135
risk evaluation, 138-139
risk indicators, 139
asymmetry coefficient, 157-159
index delta, 141-157
interrelationships between, 161-165
loss probability, 159-160
VaR (Value at Risk), 140-141
nonlinearity, options evaluation and, 1-2
nonmodal optimization, 76
nonmodal optimization space, 82
nontransitivity in multicriteria optimization, 96-97
normalization, 184

O

objective function
defined, 74
effect on optimization space, 89-91
explained, 78-79
interrelationships of, 91-96
usage of, 88
one-dimensional capital allocation system, 170-172
analysis of variance in, 190-191
factors affecting, 183-186
historical volatility in, 186-187

measuring capital concentration, 192-195
multidimensional system versus, 206-209
number of days to expiration in, 187-188
number of underlying assets in, 188-190
weight function transformation, 196-204
one-dimensional optimization, 80-82
opening signals
in delta-neutral strategy, 4
generating (in backtesting systems), 225
filtration of signals, 227-228
functionals development/evaluation, 226-227
principles of, 225-226
in partially directional strategies, 42
optimal area, defined, 75
optimal delta-neutral portfolio selection, 67-72
optimal solution
defined, 75
robustness of, 82, 102
averaging adjacent cells, 103-104
ratio of mean to standard error, 104-105
surface geometry, 106-108
optimization
adaptive optimization, 233-234
challenges and compromises in, 134
defined, 73
dimensionality of, 80-85
one-dimensional optimization, 80-82
two-dimensional optimization, 83-85
in-sample optimization, 232-233
multicriteria optimization
convolution, 97-98
nontransitivity problem, 96-97
Pareto method, 99-102
robustness of optimal solution, 102-108

steadiness of optimization space, 108-114
objective function
effect on optimization space, 89-91
explained, 78-79
interrelationships of, 91-96
usage of, 88
parametric optimization, explained, 73-75
terminology, 74-75
optimization methods
direct search methods, 115
alternating-variable ascent method, 116-118
comparison of effectiveness, 127-130
drawbacks to, 115-116
Hook-Jeeves method, 118-120
Nelder-Mead method, 123-127
Rosenbrock method, 120-123
exhaustive search, 114-115
random search, 131-133
optimization space
defined, 74
of delta-neutral strategy, 79-80
acceptable range of parameter values, 85-87
optimization dimensionality, 80-85
optimization step, 87-88
effect of objective functions on, 89-91
explained, 75-77
steadiness of, 108-109
relative to fixed parameters, 109-110
relative to historical optimization period, 112-114
relative to structural changes, 110-111
optimization step, 76, 87-88
option combinations
defined, 252
factors affecting call-to-put ratio, 44-49

long and short combinations, analysis of delta-neutrality strategy, 26-27
in partially directional strategies, 43
payoff functions for, 255
bull/bear spreads, 258-259
calendar spreads, 257-258
straddles, 256
strangles, 256-257
in portfolio
analysis of delta-neutrality strategy, 22-24
analysis of partially directional strategies, 57-59
option portfolios
capital allocation indicators
asymmetry coefficient, 180-181
delta, 180
expected profit, 179
inversely-to-the-premium, 175-176
inversely-to-the-premium versus stock-equivalency, 176-178
profit probability, 179
stock-equivalency, 174-175
VaR, 181-183
weight function for return/risk evaluation, 178-179
capital allocation systems, challenges and compromises, 214-216
features of, 169-170
multidimensional capital allocation system, 172, 204-205
one-dimensional system versus, 206-209
one-dimensional capital allocation system, 170-172
analysis of variance in, 190-191
factors affecting, 183-186
historical volatility in, 186-187
measuring capital concentration, 192-195
multidimensional capital allocation system versus, 206-209

number of days to expiration in, 187-188
number of underlying assets in, 188-190
weight function transformation, 196-204
portfolio capital allocation system, 209, 211
elemental system versus, 173-174, 211-214
option strategies, payoff functions for, 255
bull/bear spreads, 258-259
calendar spreads, 257-258
straddles, 256
strangles, 256-257
option trading strategies
limited options life span, 2-3
nonlinearity and options evaluation, 1-2
option diversity, 3
options, linear assets versus, xvii
order execution simulation (in backtesting systems), 228
commissions, 231
price modeling, 230-231
volume modeling, 229
out-of-sample testing, 232-233
out-of-the-money, defined, 252
overfitting problem, 234-236

P

parameter values, determining acceptable range of, 76, 85-87
parametric optimization, 73-75. *See also* optimization
Pareto method, 172
in multicriteria optimization, 99-102
partially directional strategies, xvii
basic form of, 42-43
call-to-put ratio, 40-42
factors affecting, 44-49
delta-neutrality applied to, 49-57
delta-neutrality strategy versus, 34

embedding forecast in strategy structure, 36-40
features of, 35
portfolio structure analysis, 57-61
payoff functions
call-to-put ratio in portfolio, 40-42
factors affecting, 44-49
defined, 253
for option combinations, 255
bull/bear spreads, 258-259
calendar spreads, 257-258
straddles, 256
strangles, 256-257
in option portfolios, 169
for separate put/call options, 254-255
percentage of profitable trades as objective function
effect on optimization space, 90
relationship with maximum drawdown, 93
relationship with profit, 92
performance evaluation indicators (in backtesting systems), 236
backtesting example, 242-245
characteristics of return, 237-238
consistency, 241
maximum drawdown, 238-239
profit/loss factor, 240-241
Sharpe coefficient, 239-240
single events, 236
unit of time frame, 236
points of delta-neutrality, 6-10
in calm versus volatile markets, 10-11, 13
quantitative characteristics of, 14-21
polymodal optimization, 76
polymodal optimization space, 82
portfolio asymmetry, analysis of partially directional strategies, 60

portfolio capital allocation approach, 173-174, 209-211
elemental system versus, 211-214
portfolio construction
capital allocation indicators
asymmetry coefficient, 180-181
delta, 180
expected profit, 179
inversely-to-the-premium, 175-176
inversely-to-the-premium versus stock-equivalency, 176-178
profit probability, 179
stock-equivalency, 174-175
VaR, 181-183
weight function for return/risk evaluation, 178-179
capital allocation systems, challenges and compromises, 214-216
classical portfolio theory, 168-169
option portfolios, features of, 169-170
multidimensional capital allocation system, 204-205
one-dimensional system versus, 206-209
one-dimensional capital allocation system
analysis of variance in, 190-191
factors affecting, 183-186
historical volatility in, 186-187
measuring capital concentration, 192-195
multidimensional system versus, 206-209
number of days to expiration in, 187-188
number of underlying assets in, 188-190
weight function transformation, 196-204
option portfolios
multidimensional capital allocation system, 172
one-dimensional capital allocation system, 170-172
portfolio versus elemental approach to capital allocation, 173-174
portfolio capital allocation system, 209-211

elemental system versus, 211-214
portfolio structure at delta-neutrality boundaries, 62-65
portfolio structure analysis
of delta-neutrality strategy, 21-34
long and short combinations, 26-27
loss probability, 31-33
number of combinations in portfolio, 22-24
number of underlying assets in portfolio, 24-25
portfolio asymmetry, 29-30
straddles and strangles, 28-29
VaR, 33-34
of partially directional strategies, 57-61
position closing signals
in delta-neutral strategy, 4
in partially directional strategies, 42
position opening signals
in delta-neutral strategy, 4
in partially directional strategies, 42
position opening/closing signals, generating (in backtesting systems), 225
filtration of signals, 227-228
functionals development/evaluation, 226-227
principles of, 225-226
premium
defined, 251
inversely-to-the-premium (capital allocation indicator), 175-176
stock-equivalency versus, 176-178
price forecasts, 35
call-to-put ratio in portfolio, 40-42
factors affecting, 44-49
delta-neutrality applied to, 49-57
embedding in strategy structure, 36-40
price modeling (in backtesting systems), 230-231

profit
concave versus convex weight function comparison, 200-202
as objective function
effect on optimization space, 89
relationship with maximum drawdown, 92
relationship with percentage of profitable trades, 92
relationship with Sharpe coefficient, 91
one-dimensional versus multidimensional capital allocation systems, 206-208
portfolio versus elemental capital allocation systems, 211-213
profit probability (capital allocation indicator), 179
profit/loss factor, 240-241
put options
defined, 251
payoff functions, 254-255

Q–R

quantitative characteristics of delta-neutrality boundaries, 14-21
quantitative characteristics analysis, 244-245
random search optimization method, 131-133
range of acceptable values, determining, 76, 85-87
rational approach to automated trading system development, xix-xx
recurrent calculations in historical database (in backtesting systems), 221-222
regression analysis, 154
reliability of data in historical database (in backtesting systems), 222-224
requirements in delta-neutral strategy, 5
restrictions in delta-neutral strategy, 5
rho, 138, 253
risk, lack of definition for, 135
risk evaluation

effectiveness of index delta, 146-149
linear financial instruments, 136-137
nonlinear financial instruments, 138-139
risk indicators, 139
asymmetry coefficient, 157-159
establishing risk management systems, 165-166
index delta, 141
analysis of effectiveness at different time horizons, 150-156
analysis of effectiveness in risk evaluation, 146-149
analytical method of calculation, 142-143
applicability of, 156-157
calculation algorithm, 141-142
example of calculation, 144-146
interrelationships between, 161
correlation analysis, 162-165
testing, 161
loss probability, 159-160
VaR (Value at Risk), 140-141
risk management
in delta-neutral strategy, 5
establishing risk indicators, 165-166
the Greeks, 169
in partially directional strategies, 43
robustness of optimal solution, 82, 102
averaging adjacent cells, 103-104
defined, 75
ratio of mean to standard error, 104-105
surface geometry, 106-108
Rosenbrock optimization method, 120-123, 129
rotating coordinates optimization method, 120-123

S

scientific approach to automated trading system development, xviii

second level of capital management systems, 167
selecting optimal delta-neutral portfolio, 67-72
selective convolution, 97
Sharpe coefficient, 239-240
    as objective function
        effect on optimization space, 90
        relationship with profit, 91
short calendar spreads, 257
short combinations
    analysis of delta-neutrality strategy, 26-27
    factors affecting call-to-put ratio, 44-49
    in portfolio, analysis of partially directional strategies, 58-59
short options, payoff functions, 254-255
short straddles, 256
short strangles, 256
signal-generation indicators in delta-neutral strategy, 4
signals generation
    in backtesting systems, 225
        filtration of signals, 227-228
        functionals development/evaluation, 226-227
        principles of, 225-226
    in delta-neutral strategy, 4
simplex search optimization method, 123-127
simulation of order execution (in backtesting systems), 228
    commissions, 231
    price modeling, 230-231
    volume modeling, 229
single events in performance evaluation, 236
slippage, 230
smoothing, advantages of, 88
smoothness of optimization space, 77
spreads, xvii
standard deviation of asset returns in risk evaluation, 136

standard error, ratio to mean, 104-105
steadiness of optimization space, 77, 108-109
relative to fixed parameters, 109-110
relative to historical optimization period, 112-114
relative to structural changes, 110-111
stock-equivalency (capital allocation indicator), 5, 174-175
inversely-to-the-premium versus, 176-178
straddles
analysis of delta-neutrality strategy, 28-29
payoff functions, 256
strangles
analysis of delta-neutrality strategy, 28-29
payoff functions, 256-257
strike price, defined, 251
strikes range index, 14-17, 51-52
structural changes, steadiness of optimization space, 110-111
structural optimization, defined, 73
surface geometry, determining robustness of optimal solution, 106-108
survival bias problem, 218
synchronization, 219
synthetic assets strategies, xvi

T

technical analysis for price forecasts, 35
testing risk indicator interrelationships, 161. *See also* backtesting systems
theta, 138, 253
three-dimensional optimization, 77
time decay, defined, 252
time horizons, effectiveness of index delta, 150-156
time value, defined, 252
transformation of weight function, 196-204
transitivity in multicriteria optimization, 96-97
two-dimensional optimization, 77, 83-85

U

unconditional optimization, 74

underlying assets in one-dimensional capital allocation system, 188-190

unimodal optimization, 76

unimodal optimization space, 82

unit of time frame in performance evaluation, 236

V

validity of data in historical database (in backtesting systems), 222-224

Value at Risk. *See* VaR

values, determining acceptable range of, 76, 85-87

VaR (Value at Risk), 181-183

- analysis of delta-neutrality strategy, 33-34
- calculation methods, 137
- drawbacks to, 140-141
- in portfolio, analysis of partially directional strategies, 61
- in risk evaluation, 136

variation coefficients, 181, 183

vega, 138-139, 253

vendors for historical database (in backtesting systems), 218

visual analysis of backtesting results, 242-244

volatile markets

- delta-neutrality boundaries in, 10-13
- delta-neutrality boundaries in partially directional strategies, 51-52
- historical volatility in risk evaluation, 136
- portfolio structure analysis
  - long and short combinations, 26-27
  - loss probability, 31-33
  - number of combinations in portfolio, 22-24
  - number of underlying assets in portfolio, 24-25
  - portfolio asymmetry, 29-30
  - straddles and strangles, 28-29

VaR, 33-34
volume modeling (in backtesting systems), 229

W–Z

walk-forward analysis, 235
weight function
for return/risk evaluation, 178-179
transformation of, 196-204

In an increasingly competitive world, it is quality of thinking that gives an edge—an idea that opens new doors, a technique that solves a problem, or an insight that simply helps make sense of it all.

We work with leading authors in the various arenas of business and finance to bring cutting-edge thinking and best-learning practices to a global market.

It is our goal to create world-class print publications and electronic products that give readers knowledge and understanding that can then be applied, whether studying or at work.

To find out more about our business products, you can visit us at www.ftpress.com.